

А. С. СМЕРНОВА

ПОСТРОЕНИЕ
АВТОМАТИЗИРОВАННЫХ
ФАКТОГРАФИЧЕСКИХ
ИНФОРМАЦИОННО-
ПОИСКОВЫХ СИСТЕМ
В ГЕОЛОГИИ

А. С. СМЕРНОВА

55,002

ПОСТРОЕНИЕ
АВТОМАТИЗИРОВАННЫХ
ФАКТОГРАФИЧЕСКИХ
ИНФОРМАЦИОННО-
ПОИСКОВЫХ СИСТЕМ
В ГЕОЛОГИИ



МОСКВА
«НЕДРА» 1976



Смирнова А. С. Построение автоматизированных фактографических информационно-поисковых систем в геологии. М., «Недра», 1976, 96 с.

Рассмотрен способ построения автоматизированных фактографических информационно-поисковых систем (АФИПС) на примере геологической информации. Описано применение АФИПС в геологии, дается их классификация и указываются средства, необходимые для создания систем каждого типа. Приведены семантические средства АФИПС. Изложены способы анализа и систематизации геологической информации, построение информационно-поискового языка и пути формализации вводимых в систему геологических понятий. Дан анализ технических и математических средств создания указанных систем.

Книга представляет интерес для геологов, занимающихся проблемами информации в геологии, а также для специалистов, разрабатывающих вопросы накопления, поиска и обработки геологических данных с применением ЭВМ.

Табл. 8, ил. 13, список лит. — 123 назв.

ПРЕДИСЛОВИЕ

Широкое применение электронно-вычислительных машин в геологии насчитывает уже свыше 15 лет. Однако это относится главным образом к использованию их в качестве быстрого вычислителя для решения различных задач с помощью математических методов. Еще очень мало используются ЭВМ для задач накопления, хранения, обработки и поиска информации на базе имеющихся массивов, т. е. для тех функций, которые выполняются в автоматизированных фактографических информационно-поисковых системах (АФИПС). Насущная потребность в создании таких систем давно не является новостью. Рост объема геологической информации, происходящий в результате увеличения глубины и масштабов исследований, числа открытых и изученных объектов, а также необходимость сохранения, анализа и переработки информации для решения различных геологических задач являются основными аргументами в пользу создания автоматизированных фактографических ИПС.

Вопросы, рассмотренные или только упомянутые в этой книге, относятся к созданию АФИПС. Они, по мнению автора, относятся к числу основных и до сих пор в достаточной мере в геологической литературе не освещены.

В I части дано общее представление об информационно-поисковых системах, их функциях, этапах построения, о состоянии дел в этой области и перспективах.

Во II части рассмотрены семантические вопросы создания ФИПС. Основное внимание здесь уделено составу геологической информации, способу представления ее в фактографических системах, правилам сравнения. Именно семантические вопросы представляют наибольшую сложность и важность в разработке многих ФИПС в геологии. Построение ФИПС рассматривается здесь применительно к собственно геологической информации, т. е. информации, относящейся к полезным ископаемым, металлогении, тектонике, петрографии, минералогии. Это наименее формализованная часть геологических данных, однако представляющая собой, по существу, основу всей геологической информации. Вопросы накопления и обработки смежной геологической информации, особенно геофизической,

представляют собой самостоятельную задачу, которая здесь не ставится.

В III части достаточно схематично рассмотрены некоторые вопросы математического и технического обеспечения ФИПС, которые должны решаться при участии специалистов в области вычислительной техники и программирования. Однако общее представление об этих вопросах должны иметь и специалисты-геологи, т. е. основные проектировщики и разработчики ФИПС, так как только с учетом всего комплекса средств создания ФИПС — информационных, математических, технических и организационных — можно проектировать систему: разрабатывать требования к ней, обуславливать режимы работы, сферу применения и т. д.

**НЕКОТОРЫЕ ОБЩИЕ ВОПРОСЫ СОЗДАНИЯ
ФАКТОГРАФИЧЕСКИХ ИНФОРМАЦИОННО-ПОИСКОВЫХ
СИСТЕМ В ГЕОЛОГИИ**

Глава 1

ФАКТОГРАФИЧЕСКИЕ ИПС И ИХ ФУНКЦИИ

Типы информационно-поисковых систем

В настоящее время известно два основных класса информационно-поисковых систем (ИПС): системы документографические (документальные) и системы фактографические. Главное отличие их между собой содержится в самом названии — первые описывают документы и служат для их поиска, вторые — факты и служат для поиска фактов. Системы документографические — это системы, заменяющие традиционные библиотечные системы поиска, представленные каталогами и указателями, а системы фактографические — это системы, заменяющие собой справочники или энциклопедии.

В последнее время поставлена задача и ведутся работы по так называемой интегрированной обработке информации [86]. В этом случае предусматривается одноразовая обработка документа, при которой из него извлекается и записывается в формализованном виде вся содержащаяся в нем информация с целью последующей выдачи сведений как о самом документе, так и о фактах, которые в нем описаны. ИПС, выполняющие подобные функции, можно отнести к третьему классу ИПС — документно-фактографическим. Однако, по-видимому, это может касаться только определенных видов документов и определенной информации, не говоря уже о тех сложностях, с которыми сопряжено создание таких систем. Попытки хранения данных вместе с их библиографией, как указывается в работе А. Хубака [111], пока являются безуспешными.

Существует мнение, что принципиальной разницы между документальными и фактографическими системами нет. Этой точки зрения обычно придерживаются некоторые разработчики документальных систем, а совершенно противоположной — разработчики систем фактографических. На наш взгляд, существенная разница между этими системами вытекает хотя бы из того факта, что они должны удовлетворять двум несовместимым в настоящее время требованиям, из которых первое — бы-

стрый и простой в использовании способ отыскания документа, а второе — возможность без обращения к документу непосредственно с помощью системы получить и обработать фактические данные. Эта разница в требованиях к системам обуславливает и различный подход к их построению, сказывающийся на способах систематизации информации и некоторых принципах построения информационно-поискового языка (ИПЯ).

Документальные и фактографические ИПС различаются, кроме того, способами ведения информационных массивов, способами обработки данных и нередко также источниками сведений. В то же время оба вида систем могут в определенной степени дополнять друг друга при проведении различных научных и производственных работ, а также в автоматизированных системах управления [2, 63].

Однако более реальным представляется путь перестройки некоторых видов документов таким образом, чтобы информацию можно было вводить с них непосредственно в память автоматизированных ФИПС. Этому следуют уже разработчики АСУ как у нас в стране, так и за рубежом при создании автоматизированных банков данных.

Термин «банк данных» (БД) или «автоматизированный банк данных» (АБД) вошел в употребление в последние несколько лет и не имеет вполне единого толкования. Иногда под ним подразумевают только массив данных, но чаще массив данных вместе с комплексом средств для поиска информации. Очевидно, следует считать этот термин синонимом термина ИПС. Однако в связи с тем, что понятие о АБД появилось недавно, чаще всего АБД приписывают особенности больших ИПС, характерные вообще для современного уровня развития в этой области, а именно — наличие больших взаимосвязанных массивов информации, пригодных для решения различного рода задач; обслуживание большого числа пользователей, в том числе удаленных от массивов, с которыми связь осуществляется с помощью дистанционных терминалов; организация массивов информации на запоминающих устройствах прямого доступа; режим работы ЭВМ диалоговый и пакетной обработки; логическая обработка информации и использование информационно-логического языка.

К ФИПС частично примыкают по своим функциям АСОД — автоматизированные системы обработки данных. В отличие от ФИПС они не накапливают информацию для долговременного хранения, а только осуществляют ее обработку. Некоторые особенности АСОД будут еще рассмотрены, но, по-видимому, они являются переходной ступенью к АФИПС, совмещающим обе взаимосвязанные функции — обработки и накопления данных.

Общими для всех ФИПС являются функции накопления и обновления информации, поиска и выдачи накопленных данных и обработки данных. Различие между ними обусловлено составом

вом хранимой информации и способом ее обработки. Понятие «обработка информации» является достаточно широким. Для того чтобы просто выдать информацию по запросу, также нужно ее обработать, т. е. осуществить поиск, сравнение, декодирование и т. п. В связи с многообразием видов обработки целесообразно различать следующие ее типы: *а* — простая семантическая обработка — сравнение на совпадение и несовпадение терминов, необходимое только для выдачи данных; *б* — сложная семантическая обработка — классификация объектов, нахождение между ними сходства и различия на основе анализа смысла информации; *в* — простая математическая обработка — суммирование, нахождение среднего арифметического, нахождение процента и т. п.; *г* — сложная математическая обработка — участие информации в решении задач математическими методами типа распознавания образов, нахождения статистических зависимостей (в отличие от обработки типа *б*, содержание отдельных элементов информации в этом случае не рассматривается); *д* — логическая обработка информации на основе логического вывода.

В зависимости от обработки, которая может выполняться в системе, имеет смысл выделить следующие типы ФИПС: 1) информационно-справочные (возможна обработка типов *а* и *в*); 2) информационно-диагностические (возможна обработка типа *г*); 3) информационно-логические (возможна обработка типов *б* и *д*).

Деление ФИПС на эти типы, хотя и опирается на известные примеры из разных областей знаний, является во многом условным. Специфика каждой фактографической системы заложена главным образом в способе представления информации (характере ее ИПЯ).

Предпосылки к созданию автоматизированных ФИПС в геологии

Необходимость создания АФИПС в геологии вытекает прежде всего из потребностей геологической практики. Важнейшей предпосылкой является огромный быстро увеличивающийся объем информации, которую требуется не только сохранить, но и как можно более интенсивно использовать.

Ежегодно геологи открывают сотни месторождений и рудопроявлений, разбуривают свыше 20 млн. м скважин, проходят свыше 300 тыс. пог. м горных выработок, выполняют около 20 млн. различных анализов (цифры взяты из книги: «50 лет советской геологии». М., «Недра», 1968), каждый из которых должен сопровождаться соответствующим геологическим описанием (из них, по данным ВСЕГЕИ [61], ежегодно выполняется свыше 400 тыс. полных силикатных анализов горных пород и минералов); описывают миллионы обнажений, опосковывают тысячи квадратных километров территории.

Все данные, полученные в результате проведенных исследований, необходимы для дальнейшего использования при поисках, разведке и оценке месторождений и территорий, для открытия и изучения новых промышленных объектов и выявления геологических закономерностей их размещения.

Геологическая информация обходится очень дорого. Ее особенность состоит в том, что во многих случаях ее нельзя повторить и только часть ее подвержена старению. Между тем сравнительно небольшой объем всей полученной информации используется в дальнейшем. В работе Г. А. Булкина и др. [45] указывается, что в области металлогении в процессе первичного обобщения (написания отчета) используется только 5—20% собранной информации, а последующее обращение к ней весьма ограничено. Аналогичная картина, по-видимому, имеет место и за рубежом, где, по данным А. Хубакса [111], количество используемой информации по месторождениям полезных ископаемых составляет менее 5% от потенциально возможной информации, находящейся в архивах, которые, как указано, становятся бесполезными. Незнание о наличии информации или трудность ее отыскания приводит к необходимости либо снова открывать открытое и изучать ранее изученное, либо обходиться минимальным числом данных. В связи с этим автоматизированные ИПС фактографического типа могут оказать большое влияние на эффективность геологических работ, повысить коэффициент использования информации и обеспечить большую достоверность выводов, для которых будут привлекаться все известные факты.

Существенной предпосылкой к созданию АФИПС является также широкое внедрение ЭВМ во многие сферы науки и производства. Ежегодно увеличивается парк ЭВМ, в том числе и в геологических организациях, совершенствуются сами машины. ЭВМ, выпускаемые промышленностью в настоящее время, могут вполне взять на себя функции активного хранилища и средства быстрого поиска и обработки информации. К вышеизложенному нужно добавить также интенсивное внедрение в геологию математических методов, которые, однако, еще не имеют значительной информационной базы для своего применения. Создание активных массивов информации в ФИПС будет способствовать еще большему совершенствованию этих методов и тем самым служить задачам геологии.

По сравнению с традиционными средствами поиска информации АФИПС обеспечивают большую эффективность за счет:

- более быстрого поиска необходимой информации;
- более полного состава данных, которые можно включить в оборот;
- более качественной и объективной обработки данных, а также получения на их основе новой синтезированной информации.

Все эти преимущества достигаются благодаря определенной систематизации накапливаемой информации, алгоритмизации и формализации процессов обработки и автоматизации самого процесса поиска данных.

Некоторые трудности на пути создания АФИПС

В отличие от автоматизированных систем поиска документов, создание которых опирается на вполне развитую теорию и достаточно четкие критерии оценки этих систем [7, 59, 88], для систем фактографических подобной теории нет. Это связано не только с большими трудностями накопления и обработки фактографической информации, чем документальной, но и с более расплывчатой функцией самих систем, с отсутствием однозначных критериев их оценки и четких требований к ним. Поэтому можно сказать, что отсутствие единой теории разработки ФИПС является одним из препятствий в их создании. Это относится только к автоматизированным ФИПС (в геологии много ФИПС, реализованных на перфокартах [39]).

Создание АФИПС не входит в задачу информационного обслуживания в его обычном понимании. Разрабатывать АФИПС могут лишь специалисты, владеющие самой информацией и знакомые к тому же с методами автоматизации поиска сведений. Таких специалистов в геологии, по-видимому, еще недостаточно, что также составляет трудность в области разработки АФИПС.

В настоящее время имеются еще некоторые технические ограничения ЭВМ (недостаточная емкость оперативной памяти, сложность ассоциативного поиска, трудоемкость ввода информации и др.), которые, по-видимому, будут преодолены в самое ближайшее время. Более существенными представляются связанные с особенностями геологического языка семантические трудности, для снятия которых нет достаточных предпосылок.

Многие геологические понятия, которыми пользуются в обычной практике, недостаточно однозначны и формализованы для применения их в автоматизированном поиске и требуют уточнения при вводе в ФИПС. Остается надеяться, что именно создание АФИПС повлияет в определенной мере на большую упорядоченность геологического языка по «принципу обратной связи» и это может быть обусловлено тем, что для ввода понятий в систему требуется их уточнение и разграничение между собой, поскольку машина не способна искать то, что не определено для нее достаточно четко.

Принципиально возможен еще один выход из семантических трудностей. Это путь диалога между машиной и человеком, в процессе которого их «взаимопонимание» должно улучшаться. Этот вариант обеспечивает решение части семантических вопросов с помощью техники. Но здесь еще также много неясного.

До сих пор не решен однозначно вопрос об источнике информации для ФИПС. Существует точка зрения, что источником информации должны быть ныне существующие опубликованные и неопубликованные документы, для которых разрабатываются ДИПС. Однако процесс извлечения из них фактов чрезвычайно трудоемок, а об автоматизации этого процесса говорить еще рано. Автоматически из документа можно извлечь только метainформацию [92], пригодную для поиска документов (например, набор значащих слов), но никак не собственно информацию об объектах. Поэтому остается либо путь извлечения фактических данных из документа человеком — такой вариант единственно возможен для ретроспективного фонда; либо путь создания новых документов, пригодных для ввода в машину без дополнительного логического анализа информации. Второй путь — формализованная запись информации на стандартных бланках используется, как уже говорилось, в АСУ и в автоматизированных системах сбора первичной информации. Возможно, в будущем стандартизация распространится и на некоторые другие виды документов.

Однако нужно сказать, что до сих пор стандартизация описаний и формализация геологических понятий, необходимая для автоматического анализа информации, еще далеко не всегда сопутствуют друг другу.

Значение АФИПС для создания АСУ и решения геологических задач

Основу автоматизированных систем управления (АСУ) составляют современные методы переработки информации для задач планирования, учета и управления. Этот принцип функционирования всех АСУ, указанный академиком В. М. Глушковым в 1969 г. [20], имеет самое непосредственное отношение к геологии, вся деятельность которой связана с получением и переработкой информации.

Понятие «геология» можно рассматривать двояко. Эта двойственность обусловлена тем, что геология, с одной стороны, является отраслью производства, а с другой стороны, областью науки. Как и любая другая наука, геология невозможна без получения и анализа информации, а как отрасль производства, где само производство состоит в добычании информации, — немыслима без переработки уже имеющихся данных. Эта особенность геологии специально подчеркивается в работе В. И. Игrevского [37], посвященной основным принципам построения АСУ-Геология. В связи с этим представляется, что АФИПС, накапливающие любой вид геологической информации, должны быть использованы в АСУ.

Помимо выделения объектов административного управления (геологоразведочные партии, экспедиции, тресты и т. д.) в АСУ

следует выделять объекты геологические и управлять их освоением. ФИПС по этим объектам должны представлять собой основную часть информационного обеспечения АСУ и использоваться как для оперативного управления отраслью, так и для решения геологических задач и прежде всего для изучения закономерностей размещения месторождений полезных ископаемых и прогнозирования новых промышленных объектов.

При этом при отборе информации в АФИПС с учетом их использования в АСУ целесообразно принимать во внимание еще два принципа, выдвинутые В. М. Глушковым [21]. Первый из них — принцип единой информационной базы — состоит в том, что при создании информационных массивов нужно исходить из всего комплекса возможных задач АСУ и, вводя информацию один раз, многократно использовать ее для решения различных задач. Отсюда, в частности, вытекает, что создаваемые в геологии ФИПС должны быть совместимы между собой в отношении обмена информацией. Второй принцип — новых задач — заключается в том, чтобы не только ориентироваться на уже известные задачи, а находить принципиально новые задачи, которые сегодня не решаются в силу того, что без АСУ невозможно переработать такую массу информации или решить задачи с большой степенью точности или с большой скоростью. В противном случае, по мнению В. М. Глушкова, не имело бы смысла строить АСУ.

По-видимому, на первых стадиях функционирования ФИПС основная роль их в АСУ будет сводиться к справочно-информационному обслуживанию и к решению отдельных задач. Но по мере накопления в них информации, совершенствования системы хранения и поиска, а также разработки соответствующих алгоритмов будет сделан переход от информационно-справочных систем к системам советующим, прогнозирующим, т.е. к информационно-логическим. Однако это окажется возможным только при соблюдении условия достоверности и однозначности вводимой информации.

Составные части АФИПС

В составе АФИПС можно различать три основные части.

1. Информационный фонд; его составляют главные информационные массивы, содержащие фактографическую информацию; вспомогательные и справочные массивы, обеспечивающие запись и поиск данных в главных массивах. К справочным массивам относится ИПЯ (состоит из нескольких частей), инструкции и правила сравнения, правила перевода на ИПЯ, система кодирования.

2. Технические средства реализации ФИПС — ЭВМ и соответствующий комплект внешних устройств для подготовки и передачи информации.

3. Математическое (программное) обеспечение (МО) ФИПС. Оно формируется из общего и специального МО.

Функционирование ИПС обеспечивается рядом самостоятельных блоков, указанных на рис. 1.

Блок подготовки информации осуществляет сбор новых данных, сбор обновлений к имеющимся данным, кодирование информации, ввод информации в систему. При этом возможны следующие варианты:

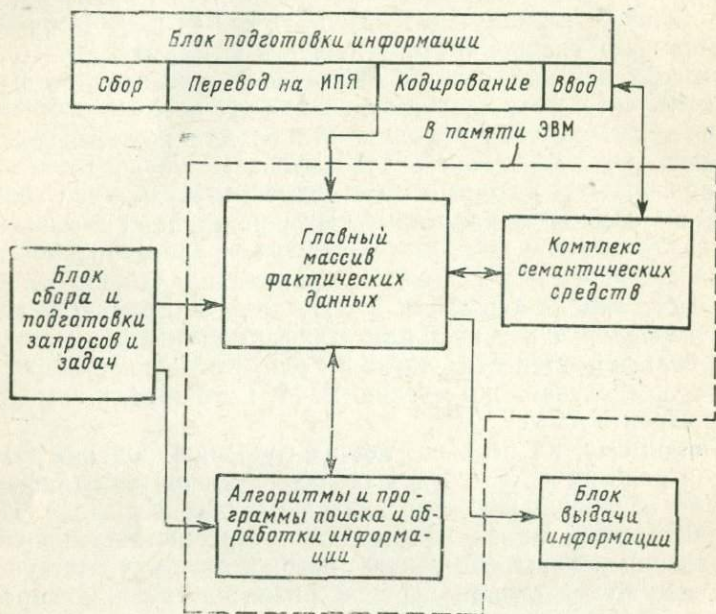


Рис. 1. Функциональная блок-схема ФИПС

1) сбор фактов и запись их на ИПЯ ведутся одновременно путем заполнения стандартных бланков с использованием словаря. Кодирование словарных характеристик осуществляется автоматически, после их ввода в ЭВМ. Целесообразен для информации, поступающей регулярно и имеющей один и тот же источник;

2) сбор фактов, их перевод на ИПЯ и кодирование осуществляется разновременно. Все операции выполняются вручную, после чего информация вводится в ЭВМ. Целесообразен для информации, извлекаемой из разнородных источников.

Блок сбора и подготовки поисковых запросов и задач осуществляет: приемку запросов, перевод их на ИПЯ, кодирование, ввод в систему. Запросы могут быть регулярные, не требующие специальной обработки, и нерегламентированные, для которых нужен индивидуальный анализ.

Главный массив (или массивы) фактических данных записан на внешних запоминающих устройствах (ЗУ) ЭВМ типа магнитной ленты или магнитных дисков (в отдельных случаях на ПК или ПЛ, которые обычно служат лишь промежуточным звеном при вводе информации). В процессе поиска и обработки информации отдельные части информационных массивов поступают в оперативное ЗУ (ОЗУ). Главный информационный массив периодически пополняется и корректируется. Он обязательно имеет дубли на магнитной ленте.

Комплекс семантических средств, в том числе информационно-поисковый язык системы записаны в ОЗУ. В случае недостаточной емкости последнего ИПЯ размещается на внешних ЗУ и по мере необходимости вводится в ОЗУ. Другой комплект средств ИПЯ имеется вне ЭВМ в составе блока подготовки информации.

Алгоритмы и программы поиска и обработки информации размещаются на ЗУ ЭВМ. Алгоритмы разрабатываются частично в составе блока подготовки задач.

Блок выдачи информации осуществляет корректировку полученных с ЭВМ данных и передачу их заказчику.

Состояние разработок АФИПС и примеры некоторых зарубежных систем

Работы по созданию автоматизированных фактографических ИПС и обработке геологической информации на ЭВМ начаты с середины 60-х годов [30, 55, 78, 103, 121]. За последние 4—5 лет число разработок значительно возросло, что позволяет выявить некоторые тенденции развития в этой области.

1. Широкое признание необходимости создания фактографических систем на базе ЭВМ и переход от перфокартных систем к автоматизированным.

2. Работа по унификации описаний геологических объектов и разработке стандартов на отдельные виды информации. Соответственно ставится вопрос о стандартизации математического и технического обеспечения этих систем.

3. Переход от локальных систем к созданию автоматизированных систем на уровне национальных (государственных) и международных.

В Советском Союзе разработкой АФИПС в настоящее время заняты многие институты и геологические управления. АФИПС предполагается создавать либо самостоятельно, либо в качестве составных частей различных АСУ, таких, как АСУ-Прогноз [29], АСУ-Поиск [28], АСУ-Вода [54], АСУ-Наука [73]. Основные работы направлены на создание систем по месторождениям полезных ископаемых [51, 70, 81], по результатам химических анализов [12], по геологии нефти и газа [3, 5, 22, 69], по скважинам, по геохимии, по полевым данным [38, 66].

Основные результаты этих работ освещены в литературе [38, 68, 69, 70]; здесь же ограничимся лишь некоторыми замечаниями.

Несмотря на широкий интерес к созданию автоматизированных ФИПС, работы в этой области находятся преимущественно на уровне исследований. Последнее обусловлено рядом объективных причин, упомянутых выше, а также слабой целенаправленностью в проведении этих работ со стороны планирующих организаций отрасли (в основном системы разрабатываются там, где есть энтузиасты этого дела); отсутствием централизованной разработки стандартов на описание всех видов геологических объектов; недостатком теоретических исследований; отсутствием полной увязки между системами в функциональном отношении, а также при разработке информационного и математического обеспечения.

За рубежом работы по созданию ФИПС (и банков данных) ведутся во многих странах, особенно интенсивно в Канаде [96, 97, 98, 101, 106, 116], США [95, 107, 122], Франции, Чехословакии [109, 112], ГДР [99], Швеции [99], Румынии [104].

С 1967 г. существует КОГЕОДАТА — комитет по хранению, автоматизированной обработке и поиску геологических данных Международного союза геологических наук. Цель этого комитета состоит в выработке рекомендаций и стандартов, касающихся геологических описаний, и распространении опыта в этой области*.

Основное направление работ во многом совпадает с проводимым в нашей стране. Наибольшее внимание уделяется созданию автоматизированных систем по нефтяным и газовым скважинам, по месторождениям полезных ископаемых, по объектам полевых исследований, по геохимии, гидрогеологии, музейным и палеонтологическим образцам и некоторым другим видам информации. Из них укажем только некоторые разработки.

В США и Канаде имеется около 2200 тыс. скважин, пробуренных при поисках на нефть и газ, около 700 тыс. из них задокументировано в двенадцати автоматизированных системах [109].

Создание банка данных по скважинам проектируется в США [95]. В нем предусмотрено хранение разнообразной геологической информации, относящейся к четырем уровням — стратиграфическим единицам, литологическим единицам, горным породам и минералам. Проект интересен способом организации информационных массивов, включением большого числа данных и полусвободным форматом записи.

* Уже вышло два библиографических указателя, включающих работы по ФИПС, проводимые в разных странах. Первый указатель охватывает период с 1946 по 1969 г. и содержит 336 ссылок (из которых 92% приходится на 1964—1969 гг.), второй — период с 1970 по 1972 г. и содержит 211 ссылок [99, 108].

С 1967 г. в Канаде на базе ЭВМ под покровительством Национального консультативного комитета по исследованиям в геологических науках создается национальная сеть сбора и обработки данных по месторождениям полезных ископаемых [101]. В 1972 г. в нее входило 10 центров, собирающих информацию по различным видам полезных ископаемых, в том числе рудным, угольным, нефтяным и газовым [97]. Общее число месторождений в Канаде оценивается в 100 тыс., из них 30 тыс. задокументировано. В 1972 г. в сеть было включено 5 тыс. рудных месторождений и 1200 залежей нефти и газа. Для рудных месторождений используется 78 входных характеристик, а для залежей нефти и газа — 543 [98]. В центрах установлены ЭВМ среднего класса, между центрами предусмотрен обмен информацией. Все центры пользуются стандартным программным обеспечением — системой САФРАС (SAFRAS — Self, Adaptive Flexible Format Retrieval and Storage System). В качестве языка программирования используется КОБОЛ. Особенностью системы является свободный формат входных документов [118].

При разработке информационной сети по месторождениям полезных ископаемых руководствуются рекомендованными комитетом КОГЕОДАТА стандартами описаний с целью перспективы совместимости данных в международном масштабе. Стандарты разрабатываются специальным подкомитетом. Они касаются различных аспектов описания месторождений, в том числе пространственной привязки [113], методов интерпретации геологических и экономических данных. Помимо систем, включенных в национальную сеть, в Канаде имеются разработки, использующие другие принципы сбора информации по месторождениям полезных ископаемых. Например, для двух провинций Канадского щита создается банк данных (БД) специально для статистического анализа с целью прогноза месторождений. БД состоит из двух массивов. Это массив геологических данных, описывающий ячейки территории размером 10×10 км, выделенных в соответствии с системой разграфки UTM, и массив данных по 500 месторождениям для Верхней провинции [106]. Всего в БД было описано 8,5 тыс. ячеек, покрывающих территорию в 85 тыс. км². Оба массива информационно увязаны.

С 1970 г. разрабатывалась и имела несколько вариантов система «Геолог». Она предназначена для сбора информации по скважинам, пробуренным в районе меднопорфировых и других месторождений [97].

Для рекогносцировочного картирования изверженных и метаморфических пород Квебекским департаментом природных ресурсов был разработан и реализован в 1968—1970 гг. так называемый Гренвильский проект, в соответствии с которым было закартировано 85 тыс. кв. миль территории и в банке данных на магнитной ленте было записано около 8 тыс. обнажений. Автоматически вычерчивалась карта обнажений, обозначались струк-

турные линии и контуры горных пород. Специальная программа, использующая данные по структуре и минеральному составу горных пород, определяла тип породы из 40 разностей [96, 123].

В США была разработана система сбора данных для картирования метаморфических и гранитных пород Аляски [122].

Теоретические вопросы разработки АФИПС и систем обработки информации рассмотрены в следующих работах [103, 105, 110, 114]. Последние три из них посвящены представлению в автоматизированных системах геологических понятий, главным образом горных пород. Идея семантического кодирования геологических понятий за рубежом разрабатывалась П. Лаффи-том (P. Laffitte) в рамках проекта Геосемантика-70.

Некоторые итоги по разработке ФИПС отражены в статьях председателя комитета КОГЕОДАТА А. Нивах [111] и члена этого комитета J. Hruška [109].

В целом можно отметить, что уровень разработок, проводимых за рубежом, достаточно различный. Большая часть АФИПС совмещает функции накопления и обработки информации. Стандартизация еще только внедряется, затрагивая не только информационное, но и математическое обеспечение систем. Создаются комплексы программ специально для обработки геологической информации. Кроме уже названной выше канадской системы SAFRAS имеются указания на систему GIPSY (Generalized Information Processing System) в США, которая используется для обработки информации по месторождениям полезных ископаемых, систему SIGMI (Système d'Information Géologique et Minière) во Франции, также применяемую для обработки информации по месторождениям полезных ископаемых, и др. [108]. В большинстве систем используются ЭВМ типа IBM/360 и IBM/370 равных моделей, имеющие ЗУ прямого доступа в виде магнитных дисков. В качестве языка программирования используются КОБОЛ, ФОРТРАН, АЛГОЛ, АССЕМБЛЕР.

**МЕТОДЫ АНАЛИЗА И СПОСОБЫ КЛАССИФИКАЦИИ
ГЕОЛОГИЧЕСКОЙ ИНФОРМАЦИИ**

Создаваемые в настоящее время фактографические ИПС, использующие дорогостоящую вычислительную технику и накапливающие не менее ценную геологическую информацию, уже только по этим признакам выходят за рамки частных локальных систем и должны отвечать ряду требований, которые позволили бы пользоваться собранной информацией широкому кругу потребителей и в течение продолжительного срока.

При проектировании ФИПС в геологии, очевидно, следует исходить из того, что любые геологические сведения могут и должны быть элементами информации в информационно-поисковой системе. Проблема, по-видимому, состоит лишь в решении следующих вопросов: какую информацию нужно вводить, а какую можно синтезировать в системе; каким образом следует систематизировать информацию, чтобы использовать ее наилучшим образом, избежать большого дублирования одной и той же информации в разных системах и сделать так, чтобы данные, накапливаемые в каждой из систем, были сопоставимы между собой.

С этих позиций рассмотрены вопросы состава геологической информации, способы ее классификации и систематизации в ФИПС и системах автоматизированной обработки данных.

Виды геологической информации и ее источники

К собственно геологической информации можно относить сведения двух основных видов:

1. Получаемую в результате полевых, лабораторных и камеральных исследований конкретную геологическую информацию, отражающую строение определенного участка (или точки) земной коры.

2. Абстрактные сведения, представляющие собой совокупность геологических знаний, зафиксированных в системе понятий. Этот вид сведений, в отличие от первых, не привязан к конкретной территории и геологической обстановке.

Информация абстрактная также должна вводиться в АФИПС, но чаще всего она в этом случае не является собственно информацией*. Ее функция состоит лишь в том, чтобы правильно записать и интерпретировать информацию о геологических объектах. Она вводится на правах языка системы, а точнее на правах ее тезауруса**. В геологии оба вида информации взаимосвязаны и взаимообусловлены.

Обобщенную схему возникновения геологической информации условно можно представить в виде следующих этапов (рис. 2):

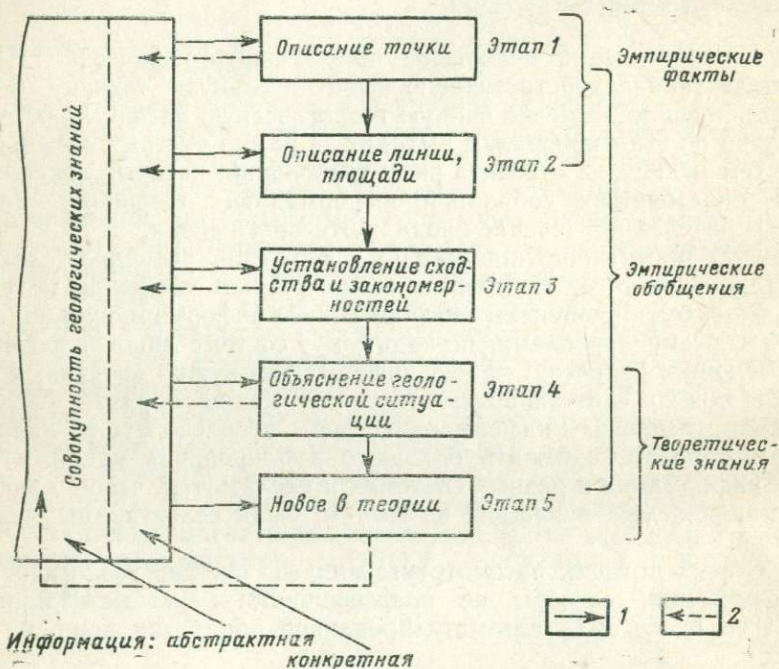


Рис. 2. Условная схема этапов возникновения геологической информации: 1 — на каждом этапе привлекается информация из общей совокупности; 2 — поступает в общий фонд

* Существуют АФИПС, в которых объектом хранения и выдачи является именно абстрактная информация. Они используются как справочники и в учебных целях. Пример такой системы — «Определитель минералов на перфокартах», разработанный в ВЗПИ [24].

** Тезаурусом принято называть систему понятий и их взаимоотношений между собой. О тезаурусе можно говорить применительно к ИПС, коллективу и т. д. [94].

Этап 1. Дискретные наблюдения о строении земной коры в отдельных точках (макро- и микронаблюдения, анализы).

Этап 2. Линейная и площадная интерполяция и экстраполяция о строении территории (построение разрезов, профилей, карт, графиков, планов, оконтуривание геологических тел).

Этап 3. Поиск аналогий со строением других участков. Изучение и выявление закономерностей нахождения геологических тел и их свойств.

Этап 4. Попытка дать объяснения условиям образования геологических тел, прогнозирование тел полезных ископаемых или других геологических образований и их свойств.

Этап 5. Дополнение, корректировка и развитие системы геологических знаний на основе нового фактического материала.

На каждом этапе используется информация, полученная на предыдущем этапе, а также информация из общей совокупности знаний. Основную информационную ценность представляет информация, полученная на 1-м этапе, так как эта информация в определенной мере невоспроизводима и на нее затрачены максимальные средства. Видимо, для нее нужно создавать АФИПСы прежде всего. Однако с точки зрения большого числа геологических задач и в первую очередь задач поиска полезных ископаемых этот уровень данных может оказаться слишком детальным. Прогнозирование месторождений и выявление закономерностей их размещения можно осуществлять только на основе обобщенной информации, появляющейся на 2-м и последующих этапах геологического изучения. В связи с этим возникает вопрос, нужно ли информацию, полученную как результат переработки первичных данных, вводить в ФИПС, или ее можно генерировать автоматически из первичной.

По-видимому, следует признать, что на данный момент генерировать автоматически всю геологическую информацию из первичной мы еще не научились, хотя частично это возможно. Например, автоматически получают различные изолинии, проводят границы геологических тел, классифицируют объекты, устанавливают между ними степень сходства, оценивают перспективность территорий.

Однако при этом используется не только первичная информация, но и обобщенные данные. Кроме того, сами математические методы, которыми пользуются геологи, нуждаются в геологической интерпретации [89], а для этого нужна геологическая информация разных уровней. При этом, по-видимому, любое геологическое описание содержит элементы обобщения уже в самом начале и не является в полной мере элементарным. Многие понятия, которыми пользуется геолог на этапе 1, сами по себе уже являются обобщениями, например названия горных пород. На этапе 2 объем используемых априорных знаний и степень обобщения фактов увеличивается. На этапе 3, кроме информации, полученной на первых двух этапах, привле-

каются данные по другим территориям, а также абстрактная информация о возможных геологических ситуациях. Этап 4 отличается тем, что здесь происходит экстраполяция не только в пространстве, но и во времени, делается попытка восстановить картину ранее действующих процессов, результаты которых наблюдаются на данном участке. Информацию, получаемую на этапах 4 и 5, непосредственно вывести из информации первого уровня нельзя — при получении этой информации используется еще некоторая неформализованная часть логического мышления, которая может оказаться не зафиксированной при первичном наблюдении.

Таким образом, по мере продвижения по этапам в процесс обработки включается все большее количество информации, в том числе и абстрактной. Для того чтобы генерировать подобную информацию с помощью ЭВМ, нужно в тезаурус системы включить весь объем имеющихся к настоящему времени геологических знаний. Алгоритмы такой переработки информации еще не созданы, хотя нет доказательств, что их создать нельзя.

Кроме того, при переходе от описаний точек к описанию территорий, т. е. от объектов сравнительно элементарных к объектам более высокого порядка, нередко используется уже другая система понятий; они становятся более высокой степени общности. Изменяется сам объект, изменяется характер описания, а также цель описания. Например, вместо понятий «минералы» и «горные породы», используемых при описании сравнительно небольших участков, для описания крупных территориальных объектов используются понятия «рудные формации» и «геологические формации», а вместо отдельных трещин и складок — тектонические структуры. Поэтому, чтобы в автоматизированной системе получить геологическое описание территориального объекта из первичной дискретной информации, нужно помимо всего прочего уметь синтезировать обобщенные понятия из более элементарных. Для этого нужна формализация иерархических отношений между соответствующими понятиями.

В связи с изложенным представляется правомерным и необходимым создание в геологии АФИПС не только по первичным геологическим данным, но и по данным, являющимся обобщением первых, тем более, что во многих первичных геологических документах (отчетах геологических партий) содержится уже обобщенный материал и в первую очередь геологические карты, описания месторождений полезных ископаемых, рудных полей и других территориальных объектов, что является не менее ценной информацией, чем полевая документация.

Вышесказанное сводится к следующему — возникновение геологической информации и анализ конкретных данных связаны с привлечением априорной информации, относящейся к другим объектам, и абстрактной; формальные способы обобщения геологической информации еще полностью не разработаны,

они могут быть математическими и логическими (синтез понятий); для решения геологических задач нужно вводить в ФИПС информацию первичную и обобщенную, относящуюся к описанию территориальных объектов.

Способы классификации геологической информации

В зависимости от задачи геологическую информацию классифицируют по различным аспектам.

Существует принцип классификации по *предметным областям* геологии. Эта классификация взята за основу при создании документальных ИПС с целью разделения документальных источников информации на самостоятельные массивы и при разработке ИПЯ для документальных систем. Во многом это деление условно, предметные классы геологии пересекаются между собой, особенно такие классы, как геохимия, минералогия, петрология, месторождения полезных ископаемых.

В фактографических системах обычно используется классификация информации по геологическим *объектам*. Она является более конкретной, чем классификация по предметным областям, хотя обе классификации пересекаются. В принципе всю геологическую информацию можно распределить по объектам.

Определение объекта в АФИПС

Под геологическим объектом понимаются участки земной коры разного размера, вычленяемые с определенной целью при геологическом изучении и описании. Из этого определения следует, что к геологическим объектам в АФИПС относятся как тела, оконтуренные геологическими границами* (собственно геологические тела), так и участки, ограниченные географическими (пространственными) границами, в том числе координатной сетью.

Можно назвать следующие основные геологические объекты, которые либо уже являются, либо могут быть объектами информации в АФИПС — пробы и образцы минералов и горных пород на различные виды анализов; зоны распространения элементов, минералов и минеральных ассоциаций; петрографические формации и тела горных пород; тектонические нарушения и другие структурные единицы; стратиграфические единицы; месторождения полезных ископаемых, рудные поля, металлогенические зоны; точки наблюдения и обнажения; скважины; площади с географическими границами.

Один и тот же геологический объект может быть предметом изучения для решения различных задач. Например, проба гор-

* Понятие о геологических границах приведены в работах Ю. А. Воронина и В. Д. Наливкина [15, 62].

ной породы может использоваться для задач как петрохимии, так и геохимии; месторождение полезного ископаемого может изучаться и как металлогенический объект, и как геохимический и т. п. Поэтому кроме выделения объекта в некоторых случаях при создании АФИПС можно выделять и *специализацию* описания объекта.

Следующим принципом классификации геологической информации может быть *способ ее получения*. Наиболее полным представляется трехступенчатое деление информации на эмпирический факт, эмпирическое обобщение, теоретическую информацию.

Под эмпирическим фактом понимается элемент информации, полученный в результате непосредственного наблюдения и изучения. К эмпирическим фактам относится, например, установление структуры и состава минерала и горной породы, установление наличия трещин определенного залегания или явления их пересечения, несогласия горных пород и т. д.

Эмпирическое обобщение — это информация, полученная в результате логического обобщения единичных эмпирических фактов. По высказыванию В. И. Вернадского [13], эмпирические обобщения не выходят за пределы установленных фактов. Граница между эмпирическим фактом и эмпирическим обобщением условна; она может проводиться по степени элементарности данных.

Теоретическая информация представляет собой объяснение эмпирических фактов и обобщений на основе гипотез. К этой категории информации относятся, например, сведения о процессах и времени образования геологических тел, об источнике вещества, о генетической связи месторождений с определенной стадией развития магматического очага или стадией развития геосинклинали. Возникновение разной по способу получения информации в определенные этапы геологического изучения показано на рис. 2. Наличие большого числа сведений, которые принадлежат к категории теоретических, является, как известно, специфичностью геологии как исторической науки.

В современной теории познания в большинстве случаев выделяется только два основных уровня абстракции — эмпирические и теоретические сведения [77, 91]. Эмпирические сведения считаются фактами науки, а теоретические — объяснениями, гипотезами. Философское толкование факта содержится в работе Л. С. Мерзона [56], в геологии этот вопрос обсуждался в работе В. Б. Ляцкого и др. [55].

Практически деление информации по способу получения связано с необходимостью установления *достоверности* геологической информации. Часть геологической информации, а именно сведения теоретического характера субъективны с гносеологической точки зрения, их достоверность подтвердить нельзя. Вот как оценивает этот вид информации Л. Бриллюэн [9, с. 10]:

«Научные теории вводят связи между эмпирическими фактами, но любая теория может быть отброшена, в то время как факты остаются, если их точно наблюдали, и эти связи сохраняются в какой-либо другой теории». Достоверность сведений эмпирического характера связана с точностью или добросовестностью проведенного наблюдения и исследования. Для оценки достоверности этого вида информации в ФИПС нет общепринятых критериев. Условно информацию можно считать достоверной, если она подтверждается несколькими исследованиями или повторными наблюдениями. Эмпирические факты и эмпирические обобщения по степени достоверности вполне сопоставимы между собой.

Следующим аспектом классификации геологической информации, который используется в современных методах исследования, главным образом при математической обработке данных, является деление ее по *степени информативности*. В результате такого деления часть признаков объектов признается неинформативной и может выбрасываться из расчетов. Обычно понятие об информативности признаков применяется в системах, решающих задачи распознавания перспективных и неперспективных объектов. Деление информации на информативную и неинформативную имеет смысл только по отношению к одной определенной задаче, так как информативность изменяется от многих причин, в том числе и от метода ее определения. Естественно, что в ФИПС, ориентированных на решение многих задач, понятие информативности при отборе информации использовать нельзя.

Геологическую информацию делят также по *форме выражения*. Обычно выделяют следующие данные: количественные (цифровые); качественные (семантические), описываемые словами; графические (например, карты) [63, 116]. Это деление в определенной степени учитывается при организации хранения и поиска информации.

Для выбора способа организации информационных массивов можно подразделять информацию также на *динамическую*, изменяющуюся со временем (например, запасы полезных ископаемых), и информацию *статичную*, не изменяющуюся за все время ее хранения (результаты анализа проб).

Требования к выбору информации для АФИПС

В разрабатываемых в настоящее время АФИПС предпочтительнее отдается систематизации информации по объектам. Этот принцип позволяет распределить и записать информацию с наименьшей избыточностью. Помимо экономии в организации информации он дает возможность накапливать в удобной форме однородные геологические сведения, которые можно использовать для статистической обработки. При описании объектов,

оконтуренных геологическими границами, учитываются следующие основные аспекты — пространственное расположение, вещественный состав, структура (строение), генезис и время образования. Для объектов, ограниченных искусственно, учитываются только первые три аспекта. Из всей информации об объектах около 80% приходится на долю вещественного состава. Важнейшей характеристикой является также структура объектов, отражающая взаимное расположение элементов, из которых состоит сам объект, т. е. являющихся, как указано Л. И. Четвериковым [89] — объектами более низкого порядка, чем рассматриваемый. К сожалению, не для всех объектов имеются унифицированные способы описания структур.

Отбор сведений для ФИПС в каждом конкретном случае связан с типом выбранного объекта, его специализацией и статусом данной системы, определяющим, в какой мере она должна сочетаться с другими. Естественно, что требования к информации в изолированных системах произвольны.

Для АФИПС, рассчитанной на обмен данными с другими системами, предлагаются следующие основные требования к отбору информации:

1) всестороннее описание объекта с учетом решения различных задач;

2) предпочтительный отбор элементарных первичных фактов и тех сведений, которые нельзя непосредственно вывести из первых;

3) стремление сократить избыточность описания за счет уменьшения в одном информационном массиве сведений, относящихся к другим объектам, т. е. сохранение уровня описания;

4) точная унифицированная пространственная привязка описываемого геологического объекта;

5) независимость состава данных от носителя информации или других технических средств;

6) проведение отбора информации в соответствии с заранее составленным перечнем возможных характеристик или стандартной анкетой. В ряде случаев целесообразно фиксировать не только положительные, но и отрицательные данные, которые необходимо в этом случае отличать от не установленных;

7) осуществление сбора информации на обычном (естественном) или формализованном геологическом языке с последующим ее кодированием. Это связано с необходимостью уменьшить первичные ошибки при передаче информации и для контроля.

Большинство перечисленных требований достаточно просты и не нуждаются в пояснениях. Поэтому остановимся только на некоторых из них.

Уровень описания объекта (см. требование 3). Целесообразно считать, что объект описан на своем «собственном» уровне, если информация относится к данному объекту, а

не к его части, или, напротив, к объекту, частью которого является рассматриваемый объект.

Примеры. Предположим, в качестве объекта описания выбрано рудное тело (РТ), тогда:

Высказывание 1: «РТ сложено минералами — пиритом, халькопиритом, кварцем» — является описанием на собственном уровне РТ.

Высказывание 2: «пирит (слагающий РТ) имеет зональную структуру и содержит примесь селена» — является описанием на уровне минералов РТ, т. е. более низком по отношению к РТ.

Высказывание 3: «примесь селена содержится в пирите в количестве 0,015%» — является описанием на еще более низком уровне.

Высказывание 4: «РТ находится в пределах антиклинали К» — является описанием на собственном уровне.

Высказывание 5: «антиклиналь К (в которой имеется РТ) — является составной частью антиклинория Н» — это описание на более высоком уровне, чем уровень рудного тела.

Различать и учитывать уровень описания нужно при разработке стандартизованного описания объекта. Это не означает, что в описание объекта не следует включать уровни выше и ниже собственных, но важно договориться, какие объекты и на каком уровне описывать. По-видимому, нецелесообразно слишком удаляться от собственного уровня выбранного объекта во избежание большого перекрытия (избыточности) в описаниях. Выбор определенного уровня описания обусловлен прежде всего тем, какие из объектов будут взяты в качестве базовых, т. е. по каким объектам будут создаваться ФИПС, рассчитанные на взаимный обмен информацией. Уровень описания может являться той границей, с помощью которой информация будет систематизироваться по объектам, так как в принципе всю геологическую информацию можно привязать к одной точке.

В качестве примера избыточного описания можно назвать вариант характеристики химического анализа пробы горной породы для АФИПС на перфокартах, приведенных в работе Л. И. Вакориной и В. А. Рудника [11]. В этот вариант описания пробы горной породы включались такие понятия: эпоха складчатости, фаза складчатости, тектоно-магматическая зона и др., что на несколько ступеней выше выбранного уровня.

В работе Е. Е. Захарова и А. С. Смирновой [33] для описания рудного месторождения в автоматизированной ФИПС в качестве самостоятельных уровней, кроме собственного уровня месторождения, взяты уровень рудного тела, уровень главных минералов рудного тела и уровень вмещающих пород. Выбор типов объектов для металлогенического изучения в соответствии с масштабом исследований приведен в работе А. А. Ивановой и Ю. И. Михайловой [36].

В число базовых объектов, для которых АФИПС должны быть созданы в первую очередь, с учетом уже имеющихся в этом направлении работ целесообразно включить месторождения полезных ископаемых, рудные тела (в том числе нефтегазо-

носные залежи и угольные пласты), вмещающие породы промышленных месторождений, главные рудные минералы промышленных месторождений, скважины глубокого заложения и опорные, пробы горных пород на химический анализ, планшеты геологических карт масштабов 1 : 200 000 и 1 : 25 000, где следует фиксировать различные геологические объекты для автоматического поиска, и другие базовые объекты. Перечень базовых объектов должен постепенно расширяться, пока вся главнейшая геологическая информация не окажется увязанной в АФИПС. Состав базовых объектов, а также аспекты и уровни их описания должны разрабатываться централизованно специальной комиссией по стандартизации геологической информации для АФИПС.

Этапы отбора информации для АФИПС

Отбор информации и накопление массивов сведений на стадии проектирования и отладки ФИПС можно представить в виде следующих укрупненных этапов, показанных на рис. 3.

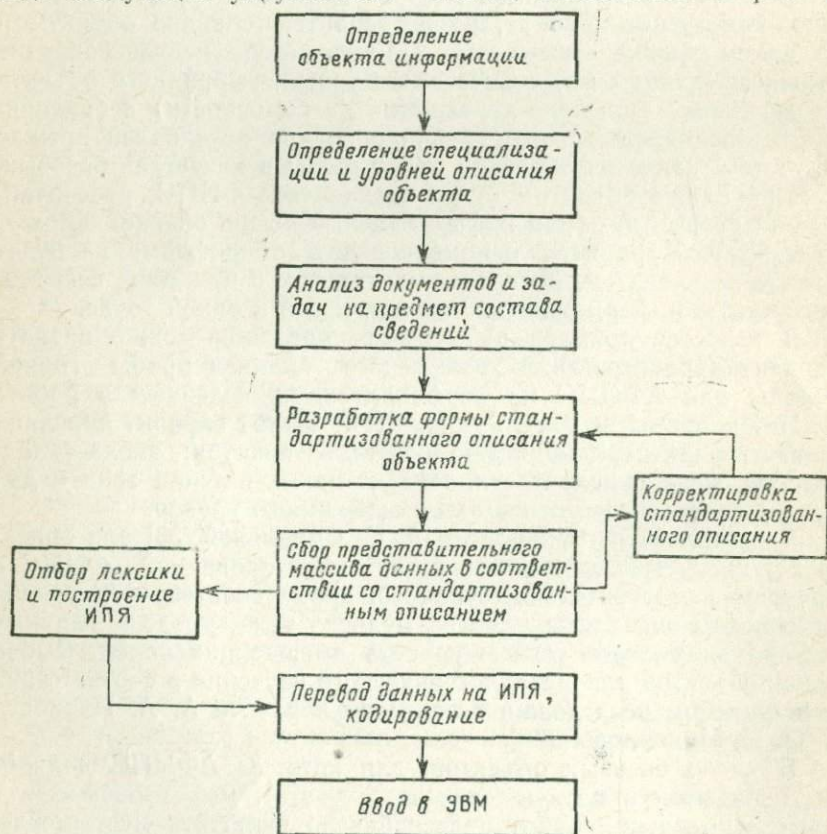


Рис. 3. Этапы отбора информации для ФИПС на начальной стадии ее организации

Обмен информацией между АФИПС

Полный обмен информацией между системами можно осуществлять только при наличии необходимых технических средств, обеспечивающих дистанционную передачу информации; совместимости математического обеспечения систем; совместимости информационной.

Информационная совместимость является важнейшей предпосылкой для обмена информацией между отдельными массивами данных и системами и позволяет решать комплексные задачи.

Целесообразно выделять два типа информационного обмена — вертикальный и горизонтальный.

Горизонтальный обмен предусматривает возможность совместной обработки информации одного уровня, относящейся к объектам одной категории. Примером обмена этого типа является возможность сопоставлять по комплексу характеристик нефтеносные залежи из разных районов страны. Такой обмен возможен при условии языковой совместимости описаний нефтеносных залежей из всех районов, т. е. на основе единого формата и единой терминологии.

К этому же виду обмена относится возможность сопоставлять одни и те же объекты, описанные в информационных массивах с разной специализацией (например, экономические сведения о месторождении и геохимические сведения о месторождении). Предпосылкой к этому является наличие одних и тех же идентификаторов объектов. Идентификатором может быть либо наименование, либо специальный код для каждого объекта, который служит для его однозначного обозначения и отличия от других объектов. По идентификатору обычно находят место хранения информации об объекте в том или ином массиве.

Вертикальный обмен предусматривает совместную обработку информации, принадлежащей разным уровням. Он обеспечивает при необходимости расширение описания объектов за счет нахождения в соответствующих массивах информации более высокого или более низкого уровня. Подобной задачей является, например, получение данных анализов по всем пробам определенного рудного тела или месторождения. Вертикальный обмен возможен, если информация разных уровней увязана между собой пространственно.

В обычной геологической практике вся геологическая информация увязывается между собой пространственно с помощью геологических карт разных масштабов. Аналогичная система увязки должна быть и для информации, записанной в машинной памяти*. Системы, выполняющие функцию увязки, должны

* В этом отношении существенное значение будут иметь ФИПС по геологическим картам [58] и ФИПС, отражающие строение территорий, ограниченных координатной сетью.

иметь вид указателей, отражающих пространственное вхождение объектов мелких в объекты более крупные, их идентификационные номера и место хранения информации о них. Примерная форма такого указателя приведена в табл. 1.

Таблица 1

Указатель нахождения геологических объектов в пределах более крупных объектов

Основной объект			Подчиненный объект			
наименование	идентификатор	идентификатор массива	наименование	идентификатор	отношение к основному объекту	идентификатор массива
Бакальское месторождение железорудное	М 14301	ИМ 06	Проба на химический анализ	ХА 4035164	Вмещающая порода	ИМ 09
				ХА 4035165		ИМ 09
				ХА 4035166		ИМ 09
				и т. д.		
	М 14301	ИМ 06	Проба на спектральный анализ	СА К 7814	Рудное тело	ИМ 17
				СА К 7815		ИМ 17
				и т. д.		

Вход в этот массив должен быть по всем названным данным, особенно важен вход по идентификаторам.

Таблица 2

Указатель принадлежности геологических объектов к территориальным единицам

Номенклатура листа	Включенные объекты		
	наименование	идентификатор	идентификатор массива
Р-41-XXI	Месторождение К ₁	М 08527	ИМ 06
	Месторождение К ₂	М 08528	ИМ 06
	⋮		
Р-41-XXI	Скважина	Скв. 1305	ИМ 08
	⋮	Скв. 1308	ИМ 08
	⋮		
Р-41-XXI	Нарушение	Т*491	ИМ 91
	⋮		
	и т. д.		
Р-41-XXII	Месторождение А ₁		
	⋮		
	и т. д.		

Аналогичные указатели целесообразно составлять на определенные территории. В них должны указываться типы объектов, находящихся на этой территории, их идентификаторы и место хранения информации о них в виде идентификатора массива (табл. 2). Состав сведений в указателях можно расширять за счет координат каждого объекта и других характеристик; тогда в главном массиве объекты будут представлены их идентификаторами и комплексом оставшихся сведений. При организации информационных массивов для указателей целесообразно использовать списки и списковые структуры, рассмотренные в главе 6.

В целом увязка всей геологической информации в единый автоматизированный информационный фонд геологических данных обеспечивается выполнением следующих условий:

- выделением базовых объектов, относительно которых должна концентрироваться информация в АФИПС;
- составлением унифицированных и стандартизованных описаний для этих базовых объектов;
- созданием ИПЯ для описания каждой категории базовых объектов;
- созданием языковых средств, обеспечивающих совместимость между собой описаний объектов разных уровней;
- разработкой унифицированной системы пространственной привязки для объектов каждого уровня;
- разработкой унифицированных идентификаторов объектов;
- разработкой специальных справочных систем (каталогов), в которых фиксируется соотношение объектов разных уровней.

Глава 3

ИНФОРМАЦИОННО-ПОИСКОВЫЙ ЯЗЫК ФАКТОГРАФИЧЕСКИХ СИСТЕМ

Общие сведения о ИПЯ и формах представления фактографической информации

Информационно-поисковый язык (ИПЯ) представляет собой комплекс средств, с помощью которых информация записывается в ИПС и находится по алгоритмам. Основные характеристики ИПЯ, по которым их можно сравнивать, следующие:

- 1) семантическая сила, т. е. способность записать любое высказывание, любую информацию;
- 2) качество поиска — полнота, точность и однозначность выданной информации;
- 3) простота обращения к информации, записанной на данном ИПЯ, и степень сложности перевода с естественного языка (ЕЯ) на ИПЯ.

Из всех языков максимальной семантической силой обладает ЕЯ, т. е. язык, на котором происходит общение в обычных условиях. Поэтому наиболее совершенной с точки зрения первого критерия была бы система, в которой всю информацию можно было бы записывать на ЕЯ и выдавать по любым запросам. Отпала бы необходимость в переводе с ЕЯ на ИПЯ. В то же время и именно в силу своих богатых семантических возможностей ЕЯ обладает наибольшей неоднозначностью, и обмен информацией между потребителем и системой невозможен (так как не удовлетворяется второй критерий).

Поэтому ИПЯ, которые используются в ИПС — это упрощенные модели ЕЯ, в которых сокращены и формализованы средства передачи смысла настолько, чтобы можно было автоматизировать процесс сравнения и поиска информации.

В создании информационно-поисковых языков имеется два разных пути передачи смысла информации, определяющие структуру записи.

Первый путь предполагает использование для записи информации готовых фрагментов текста естественного языка. Такой способ обычно используется в автоматизированных системах обработки данных — АСОД. Он рассмотрен ниже.

Второй путь передачи информации — моделирование языка, т. е. построение высказываний из отдельных смысловых элементов ЕЯ. Это направление в создании ИПЯ для АФИПС обеспечивает разработку языков высшего типа — информационно-логических для ИЛС, в которых имеется возможность не только выдавать сведения, но и синтезировать новую информацию, а также осуществлять семантический анализ записанных данных.

Представление информации в АСОД

В автоматизированных системах обработки данных участвуют обычно те же самые свойства геологических объектов, что и в ФИПС. Различие имеется в форме записи информации. Как правило, сведения об одном объекте в АСОД представлены после кодирования в виде многоразрядного двоичного числа (M).

$M = a_1, a_2, a_3, \dots, a_n$, где $a_i = \{0, 1\}$ и соответствует некоторому признаку объекта. Весь набор бинарных признаков составляет информационный язык данной системы. Для него характерно то, что отношения между единицами языка (признаками) отсутствуют и каждая из них представлена фрагментом текста на ЕЯ.

Пример. В работе Л. Н. Дуденко [27] для описания сульфидного медно-никелевого месторождения использованы следующие признаки:

- в составе интрузивного массива есть габбро-нориты;
- в составе интрузивного тела есть пироксениты (аналогично перечисляются и другие разности пород);
- в составе вмещающих пород есть вулканогенные;

- в составе вмещающих пород есть осадочные;
 - в составе вмещающих пород есть магматические;
 - метаморфизм вмещающих пород в гранулитовой фации;
 - метаморфизм вмещающих пород в фации зеленых сланцев и т. д.
- Перечисляется еще несколько фаций. Таким же путем составлены признаки о возрасте интрузивного тела и др.

Аналогичным образом описаны объекты в других АСОД [10, 46, 116]. В работе Ф. П. Кренделева и др. [46] среди признаков, описывающих золотоносные конгломераты, можно назвать следующие:

- в подстилающей толще известны основные эффузии и интрузии;
- в основании толщи имеется горизонт грубообломочных конгломератов;
- в составе аксессуарных нет магнетита.

Из приведенных примеров ясно, что набор признаков в АСОД ориентирован на описание объектов узкого класса. Отношения между единицами языка отсутствуют. При таком представлении информации система может ответить только на такие вопросы:

Например: «Имеется в составе аксессуарных магнетит?». Ответ был бы либо «да», либо «нет». На вопрос: «Какие аксессуарные вообще имеются (или не имеются)?» система ответить не могла бы.

Использовать язык рассмотренного типа в качестве ИПЯ неудобно, так как для описания объектов широкого класса потребовалось бы иметь тысячи признаков, составленных подобным образом. Язык этот очень негибок как для записи информации, так и для формулирования запросов (не удовлетворяются критерии 1 и 3).

Более удобным способом записи фактографической информации являются объектно-характеристические таблицы, которые используются как в АСОД, так и в ФИПС.

Объектно-характеристические таблицы (ОХТ)

Запись информации об одном объекте (M_i), представляющая собой строку в ОХТ, имеет вид

$$M_i = a_i^1, a_i^2, a_i^3, \dots, a_i^n,$$

где $a_i^j \in A^j$; $\{a_i^j\}$ — множество значений характеристики A_j ; $A^j \in A$; A — множество характеристик объекта.

ОХТ составляется по форме, указанной в табл. 3. В данном случае языком описания объектов являются перечень названий признаков $\{A^j\}$ и перечень значений признаков $\{a_i^j\}$. ОХТ удобно использовать для характеристик, имеющих количественные значения, например для результатов анализов горных пород и минералов, и для качественных взаимоисключающих признаков, имеющих по одному значению каждого признака на объект. Обычно значения признаков, как и в первом варианте

представления фактографической информации, не связаны между собой смысловыми отношениями, хотя сама форма ОХТ не исключает возможности использования этих отношений*. Запросы к информации, представленной в виде ОХТ, могут быть в следующих вариантах: 1) назвать значение признака A^j для объекта K ; 2) назвать все значения признаков для объекта K ; 3) назвать номера объектов, имеющих значение признака a_i^j .

Запись информации в виде ОХТ представляет собой переход от системы бинарных признаков к более свободной записи с использованием информационно-поискового языка. Наличие ИПЯ позволяет организовать менее жесткую систему записи, необходимую главным образом для описания сложных геологических объектов, каждая характеристика которых может быть представлена набором смысловых значений, и для записи информации в виде текста, т. е. во всех тех случаях, когда форма записи в виде ОХТ является слишком тесной.

Роль ИПЯ для поиска информации. Модель понимания

Для того чтобы обеспечить однозначный поиск информации на запросы, поставленные на ЕЯ, нужно эксплицировать и ввести формально в состав ИПЯ то скрытое содержание, которое в ЕЯ имеется и непосредственно в тексте не записывается. Без этой априорной информации невозможно отождествлять сведения, записанные разными словами, но имеющими один и тот же смысл. Поэтому построение ИПЯ должно опираться на модель понимания, что позволило бы обеспечить однозначное представление смысла информации.

Полное понимание между двумя системами** обеспечивается следующими условиями:

- 1) совпадением исходного множества понятий каждой из систем;
- 2) совпадением содержания каждого из понятий;
- 3) однозначным пониманием сочетаний понятий, образующих сложные понятия и высказывания.

Неполное понимание, соответствующее неполному обмену информацией, имеет место, если выполняются не все условия или часть условий выполняется неполностью. Отсутствие понимания, а следовательно, отсутствие обмена происходит при несовпадении всех названных условий. В зависимости от выполнения тех или иных условий понимание обычно рассматривается на двух основных уровнях — на уровне понятий (терминов) оно обеспечивается выполнением первого и второго условий, а на

* Жесткое распределение значений признаков ОХТ можно интерпретировать как ИПЯ с позиционной грамматикой.

** Одна из систем — это АФИПС, а другая — потребитель, посылающий запрос в АФИПС.

Форма объектно-характеристической таблицы

Наименование объектов	Наименование характеристик			
	A^1	A^2	...	A^n
Объект 1	a_1^1	a_1^2		a_1^n
Объект 2	a_2^1	a_2^2		a_2^n
Объект 3	a_3^1	a_3^2		a_3^n
	⋮	⋮	⋮	⋮
Объект m	a_m^1	a_m^2		a_m^n

уровне высказываний (текста) — выполнением всех трех условий.

Пример. Возьмем два различных текста.

Высказывание 1. «Месторождение приурочено к измененным гранитам варисского возраста и представлено вкрапленностью сульфидов молибдена и меди».

Высказывание 2. «Вторичные кварциты, развитые по гранит-порфирам, включают халькопирит и молибденит промышленного содержания. Гранит-порфиры относятся к герцинскому возрасту».

Оба приведенных высказывания описывают сходные между собой геологические ситуации, но констатировать это сходство могут лишь системы, обладающие определенным запасом априорных знаний. В частности, варисский и герцинский — эти термины обозначают одно и то же понятие*, а гранитоид и гранит-порфир — понятия, между которыми имеется отношение совместности: сульфиды молибдена и меди в высказывании 1 и халькопирит и молибденит в высказывании 2 обозначают близкие понятия (отношение род — вид).

Полное понимание текста высказываний 1 и 2 означало бы оценку данной геологической ситуации скорее всего как месторождения медно-порфирового типа. Таким путем на основе ряда фактов можно прийти к синтезу нового понятия, т.е. понятия более высокого уровня обобщения, каким является в данном случае понятие типа месторождения.

Таким образом, ИПЯ можно характеризовать в функциональном отношении как средство обеспечения ФИПС комплексом знаний на уровне тезауруса и как средство записи фактических сведений.

* Причем только в некоторых текстах. Правильно считать эти понятия совпадающими частично [19].

Составные части ИПЯ

ИПЯ, являясь моделью естественного языка, должен содержать его элементы, позволяющие передавать смысл информации. Адекватность этой модели, как и любой другой, может быть различной. Теоретически коэффициент адекватности ($K_{ад}$) можно выразить формулой

$$K_{ад} = \frac{\text{множество смыслов, которые можно записать на ИПЯ}}{\text{множество всех смыслов на ЕЯ}}; K_{ад} \leq 1.$$

$K_{ад}$ обусловлен назначением системы. Минимальным $K_{ад}$ обладает ИПЯ, состоящий только из лексики. $K_{ад}$ максимальный у ИПЯ, включающего всю лексику ЕЯ и моделирующего два типа отношений между единицами ЕЯ: 1) лексических, соответствующих отношению между понятиями (в ИПЯ они называются парадигматическими) и 2) контекстуальных (ситуативных), называемых в ИПЯ синтагматическими [59]. Кроме набора средств передачи смысла в информационных языках, можно говорить также о полноте и точности передачи смысла каждым из средств, которые соответственно могут быть различными. Например, $K_{ад}$ на уровне документальных ИПС соответствует модели, включающей, как правило, лексику и небольшую часть парадигматических отношений с точностью (в дескрипторных языках) первого приближения. $K_{ад}$ информационно-логических систем должен быть максимальным.

Таким образом, ИПЯ, являющийся наиболее адекватной моделью ЕЯ, состоит из трех частей: лексики, парадигматических отношений, синтагматических отношений. При этом каждая из составных частей моделирует соответствующие отношения ЕЯ с максимальной степенью полноты и точности.

Рассмотрим составные части языка и способы их задания.

Лексика в ИПЯ задается двумя способами — списком всех терминов языка или списком всех элементарных терминов и правил, по которым образуются сложные термины. Списки могут быть разными — классификационные кодовые словари, дескрипторные словари и просто перечни употребляемых терминов.

Парадигматические отношения в определенной мере моделируют отношения между понятиями*. Существуют следующие способы задания этих отношений в ИПЯ:

- в виде специальных отметок рядом с термином;
- графическим способом в виде классификационных схем;
- последовательным кодированием, отражающим иерархию понятий;
- семантическим кодированием.

* Для языков документальных ИПС используется часто и другой принцип установления парадигматических отношений, основанный на прагматических связях между терминами.

Семантическим кодированием называется представление на ИПЯ сложного понятия как совокупности элементарных смыслов, называемых семантическими множителями [93, 119].

Таблица 4

Виды отношений между понятиями и терминами ИПЯ

Виды отношений между понятиями	Примеры	Парадигматические отношения между терминами ИПЯ АФИПС
Совместность		
Тождество	Долеритовая структура — диабазовая структура	Синонимия
Пересечение	Медные минералы—сульфиды	Априорно не устанавливаются
Подчинение	Окислы железа — гематит	Подчинение
Несовместность		
Противоположность	Пологое залегание — крутое залегание	Несовместность
Противоречие	Измененный — неизмененный	Несовместность
Соподчинение	Девон — карбон	Частичное сходство или несовместность (в разных задачах)
	Рассланцевание — катаклаз	Совместность

В табл. 4 в качестве примера указаны виды отношений между понятиями по их объему, известные из логики [42]. Им сопоставлены парадигматические отношения между терминами ИПЯ АФИПС, описанные Е. Е. Захаровым и А. С. Смирновой [31], которые моделируют данные отношения между понятиями.

Как видно из табл. 4, большинство видов отношений между понятиями в данном ИПЯ моделируется. Неоднозначность моделирования отношения соподчинения связана с тем, что само это отношение не является достаточно четким, так как для многих пар понятий всегда найдется некоторое третье, которое по отношению к ним является вышестоящим (родовым).

Кроме того, парадигматическими связями ИПЯ моделируются отношения между понятиями по их содержанию. Например, понятия «галенит» и «церуссит» по содержанию имеют общий

признак — наличие Pb; для таких понятий в языке устанавливаются отношения частичного сходства [31].

Синтагматические отношения в ИПЯ. Известно четыре наиболее распространенных способа выражения синтагматических отношений в ИПЯ, которые используются одновременно, взаимно дополняя друг друга в одном ИПЯ, а также независимо в разных системах. К ним относятся указатели связи, указатели роли, синтагмы и стандартные фразы. Стандартная фраза является многоместным отношением (многоместным предикатом) и представляет собой форму для высказываний, в которой число мест фиксировано и каждое место функционально определено. Заполненная конкретными значениями форма является сообщением на ИПЯ, которое переводится на ЕЯ в виде грамматически законченной фразы. Стандартные фразы могут быть различной сложности. Сложные стандартные фразы включают в себя ряд простых, разделенных знаками пунктуации.

В следующих двух разделах способы построения ИПЯ для АФИПС рассмотрены более детально, сначала на уровне лексики, а затем на уровне высказываний.

Способы формализации геологических понятий.

Парадигматика ИПЯ

К числу основных категорий геологических понятий, которые участвуют в описании рассмотренных выше геологических объектов, относятся следующие: элементы, минералы, горные породы, стратиграфические подразделения, тектонические формы, формы залегания горных пород, рудных тел, минералов и другие геологические образования, процессы, свойства минералов, горных пород и веществ (по составу — структуре, элементам залегания, характеру изменения, возрасту, форме). Кроме того, существенная роль принадлежит названиям геологических и рудных формаций, типам руд и типам месторождений и другим понятиям.

Рассмотрим способы формализации и способы кодирования для ИПЯ ФИПС основных из перечисленных понятий.

Под формализацией для целей информационного поиска понимается определение понятия через его существенные признаки и эксплицитное представление этих признаков в ИПЯ. Существенными признаками понятия или предмета, следуя общей логике [42], считаются те признаки, которые являются необходимыми и достаточными для однозначной идентификации каждого из понятий (объектов).

В ИПЯ возможна неполная экспликация содержания понятия в тех случаях, когда полную формализацию провести нельзя в силу объективных причин (отсутствие определений, неоднозначность понятий и т. д.) и в тех случаях, когда для целей информационного поиска нужна только часть содержания по-

нения. В то же время очевидно, что степень формализации (полная или ограниченная) и вытекающие из этого отношения между единицами языка влияют на семантические возможности ФИПС — ее гибкость, способность адаптации к разным запросам, способность к анализу и синтезу информации.

Кодирование информации представляет собой условную замену одного обозначения другим, более коротким с целью экономии места и удобства использования. Чаще код бывает числовым, реже буквенным. Кодирование понятий и представление их на ИПЯ не исключают друг друга.

Минералы

Для минералов известны следующие формы представления в разных ИПС.

1. Перечисление без указания парадигматических связей.
2. Перечисление с мнемоническим кодированием по названию.
3. Дескрипторное.
4. Иерархическая классификация в сочетании с последовательным кодированием (аналогично УДК).
5. Смысловое кодирование по фасетной классификации Рангатана.
6. Смысловое кодирование по химической формуле.

Представление минералов без указания парадигматических связей. Используется главным образом в неавтоматизированных системах, а также в автоматизированных, но с небольшим словарем. Понятия в этом случае между собой автоматически не сравниваются.

Мнемоническое кодирование используется в некоторых системах со сбором первичной информации в поле [97, 121]. Примеры кодирования: Ас — актинолит, Аb — альбит, Рх — пироксен, Ру — рутил. Само по себе такое кодирование не является информационно-поисковым языком и так же, как и первый способ, не позволяет сравнивать отдельные минералы алгоритмически. Система отношений между ними должна вводиться в ЭВМ самостоятельно.

В дескрипторных словарях минералы представлены по их названиям с указанием некоторых видов парадигматических связей между терминами. Связи отражают в основном принадлежность минералов к классам по химическому составу или по использованию. Для фиксации полного содержания этих понятий потребовалось бы ввести большое количество отсылок к смежным терминам, что значительно усложняет процедуру поиска.

В УДК в коде минерала раскрывается вся система иерархических отношений. Например, «кварц» в этой системе имеет индекс 549.514.51 [85], в котором составные части означают

соответственно: 549 — минералогия, 549.51 — окислы, 549.514 — окислы четырехвалентных элементов, 549.514.5 — окислы кремнезема.

В фасетной классификации Ранганатана [120] кодирование минералов происходит по тем же правилам, что и химических соединений, в зависимости от положения составляющих их элементов в периодической системе. Элементы кодируются так: Cu — E114, S — E163, O — E162, где E1 — условный код для всех элементов, а числа 14, 63, 62 обозначают соответственно группу элемента и ряд, в котором они находятся. Для окислов и гидрокислов условным кодом является E2, для солей — E4. Например, E2434 — это код SiO_2 . Здесь 43 — код кремния (по периодической системе), а 4 — его валентность в данном соединении.

Смысловое кодирование. По сравнению со способами, перечисленными выше, более полная экспликация понятий минералов достигается использованием смыслового кода, предложенного А. С. Смирновой [78]. Смысловой (семантический) код в данном случае представляет собой запись минералов по их химической формуле с указанием особенностей структуры. Он позволяет однозначно идентифицировать каждый из минералов, в том числе минералы, имеющие одинаковый химический состав, но разную структуру. Например:

сфалерит $\text{ZnS}\alpha_1\beta_1$,
вюрцит $\text{ZnS}\alpha_2\beta_2$,

марказит $\text{FeS}_2\alpha_5\beta_4$,
пирит $\text{FeS}_2\alpha_1\beta_4$.

Здесь α_1 — кубическая сингония, α_2 — гексагональная сингония, α_5 — ромбическая сингония, β_1 — координационный структурный мотив, β_2 — каркасный структурный мотив, β_4 — островной [67]. Основная масса минералов представляется без структурных характеристик только по химической формуле. Отношения между минералами специально не фиксируются, так как они устанавливаются алгоритмически путем сопоставления смысловых кодов. Например, при сравнении формул германита — $\text{Cu}_3(\text{Fe}, \text{Ge})\text{S}_4$ и борнита — Cu_5FeS_4 алгоритмически на основе анализа формул можно установить принадлежность их к одному классу сульфидов и принадлежность к группам минералов, содержащих железо и содержащих медь.

Таким образом, запись минералов по химической формуле является формальным языком в отличие от записей минералов по их названиям и эксплицирует основное содержание этих понятий. Очевидно, что язык химических формул в области автоматизированного поиска данных в геологии найдет такое же широкое применение, какое он находит в химии.

Современные ЭВМ имеют достаточно развитый алфавит, позволяющий вводить химические формулы без замены латинских букв. При этом формулы линейризируются, т. е. записывается не Fe_3O_4 , а Fe3O4 . Минералы простого состава (окислы,

сульфосоли, карбонаты, сульфаты) удобнее записывать эмпирическими формулами, а не структурными. Силикаты, напротив, представляются в виде структурных формул, которые более общеприняты и глубже отражают структуру минералов.

Одним из преимуществ семантического кодирования является возможность записывать данные о минералах на любом уровне детальности в зависимости от состава семантических множителей так, что все уровни оказываются сопоставимыми между собой. Это достигается тем, что для записи классов минералов или группы используются те же символы, что и при записи полных химических формул. Примеры сокращенных записей: карбонаты- (CO_3) , силикаты-Si, окислы железа-(Fe) (O), сульфиды меди и цинка- $(Cu+Zn)S$, пироксены- (Si_2O_6) , амфиболы- (Si_4O_{11}) . Перед каждой такой записью стоит дефис, указывающий, что формула сокращена.

Семантическое кодирование в ИПС должно сочетаться с условным цифровым кодированием для того, чтобы при записи сведений о конкретных объектах экономить место в памяти ЭВМ. Тезаурус ИПС обязательно должен содержать полный семантический код, где каждому коду сопоставлено наименование минерала и условный код.

Горные породы

Горные породы как основные «носители геологических явлений» [83] участвуют при описании большинства геологических объектов. Представление их на ИПЯ, как и минералов, различно и обусловлено особенностями каждой ИПС. Основные из известных способов представления горных пород на ИПЯ указаны в табл. 5.

Первые четыре способа не отличаются принципиально от таковых для минералов. Использование в автоматизированной фактографической системе иерархической классификации горных пород указано в работе Корнелиуса и Бурка [102]. Понятия представлены здесь следующим образом:

Изверженные породы	1000
Граниты	1A00
Кварцевые фельзиты	1AB1

Семантическое кодирование по П. Лаффиту

По П. Лаффиту, каждый термин, определяющий горную породу, представлен n -битами, где n — общее количество возможных характеристик для описания однородной совокупности горных пород. К числу характеристик, на которые дается либо утвердительный, либо отрицательный ответ, относятся, например, такие: a_1 — структура порфировая, a_2 — содержит кварц, a_3 — содержит полевой шпат, a_4 — глубинная, a_5 — цеолитов более 10%, a_6 — кальция по данным химического анализа более 5% и т. п. Семантическое кодирование из A горных пород мож-

Виды представления горных пород в различных ИПС

Способ записи горных пород	Область применения	Основные преимущества	Основные недостатки с позиций АФИПС
1. Кодовые словари наименований	ИПС на ПК и АСОД	Простота кодирования	Горные породы несопоставимы между собой
2. Дескрипторное (по наименованию)	ДИПС и ФИПС	Фиксация отношений между терминами	Частичная экспликация понятий
3. Иерархические классификации с последовательным кодированием	ДИПС и ФИПС	Взаимоотношения понятий указаны в самом коде	Одноаспектное сравнение понятий
4. Мнемоническое кодирование наименований	ФИПС	Легкость запоминания кода	Кодирование не заменяет ИПЯ, понятия несопоставимы
5. Семантическое кодирование по Лаффиту и по Диксону	АФИПС (ИЛС предположительно)	Многоаспектность описания горной породы, возможность сопоставлять ее свойства	Неудобная форма записи и др.
6. Семантическое кодирование по автору	АФИПС и ИЛС	Аналитическая форма записи, сопоставимость с другими характеристиками вещественного состава и др.	Непривычность записи

но осуществить с помощью матрицы $N \cdot n$, где N — число терминов, которые используются для наименования этих пород. Если окажется, что одна и та же последовательность битов указана для двух горных пород, то вводится новая характеристика, которая позволит их различать. Для некоторых горных пород определенные вопросы характеристики не имеют смысла, поэтому фактически в предложенном способе для каждого вопроса отводится два бита, из которых первый является битом соответствия. Только в случае единицы в бите соответствия имеет смысл ответ на данный вопрос. Для записи одной горной породы используется обычно одна строка 80-колонной перфокарты

[115]. К числу недостатков в предложенном способе кодирования горных пород можно отнести следующие:

1) форму записи, которая не может существовать отдельно от носителя информации или списка вопросов, что неудобно при описании геологических объектов, где горная порода является лишь одной из многих характеристик;

2) неформализованность характеристик горных пород, среди которых есть признаки, существенные для определения типа горной породы и несущественные (например, цвет).

П. Лаффит считает, что формализация понятий должна не предшествовать семантическому кодированию, а быть его следствием. Это положение представляется ошибочным, так как в этом случае трудно будет построить язык из-за возможных дискуссий о том, какие признаки включать, а какие — нет.

Похожий способ семантического кодирования только для обломочных пород указан в работе А. Л. Димитриу [104]. Здесь каждая горная порода представлена девятью битами. Каждый бит соответствует определенному признаку. Признаки даны по размеру составляющих частиц: 1 — валун, 2 — булыжник, 3 — галька, 4 — крупный песок, 5 — средний песок, 6 — мелкий песок, 7 — алеврит, 8 — ил, 9 — глина. Примеры кодирования понятий:

аргиллит	000 000 111
глина	000 000 001
валунная глина	100 000 001
обломочная порода	111 111 111
конгломерат	011 000 000

Семантическое кодирование горных пород по семантическим множителям

Этот способ был предложен автором настоящей работы [78]. В основу метода положен тот же принцип, что и для представления на ИПЯ минералов, т. е. с выделением существенных признаков (семантических множителей), которые полностью эксплицируют понятие, обозначенное тем или иным термином, и построение формул для каждой горной породы по аналогии с химическими формулами для минералов. Построение семантического кода опирается на определение горной породы как парагенетической ассоциации минералов с определенной структурой.

Для построения семантического кода любой горной породы взяты следующие группы семантических множителей:

I. Показатели структуры горных пород (40 наименований).

II. Породообразующие минералы (150 наименований).

III. Количественные и качественные определители минерального состава (7 определителей).

Перечислим здесь только последние семь определителей.

- А — минерал составляет подавляющую часть породы ($>90\%$).
 Б — минерал преобладает над другими ($50-90\%$).
 В — минерал в числе основных породообразующих ($5-50\%$).
 Г — минерал в числе характерных, но количество его мало ($<5\%$).
 Д — минерал присутствует, но не определяет породу.
 Е — присутствует в виде вкрапленников или обломков.
 Ж — минерал составляет цемент или базис породы.

Комбинации из разных семантических множителей позволяют описать любую горную породу однозначно и на разном уровне детальности. В числе названий структур имеются общие понятия: «кристаллизационная», «обломочная», «пирокластическая», «бластическая». С их помощью обозначается название классов горных пород. Например,

- С₁ — «порода магматическая»,
 С₂ — «порода интрузивная».

Для однозначного обозначения наименования мономинеральной породы достаточным является указание ее структуры и минерала с определителем А. Например, С₃₂А₈₁ — обозначение кварцита (здесь С₃₂ — гранобластовая структура, 81 — порядковый номер кварца в списке породообразующих минералов). Если данную горную породу нужно описать более детально, то кодовая формула дополняется другими определителями. Так, С₃₂А₈₁Г₅₆ — означает присутствие в кварците характерной примеси серицита (№ 56). Для полиминеральных пород используются определители Б и В. Например,

- С₃Б₉₂В_{87,81}Г₅₄₍₃₆₎ — кодовая формула гранита,
 С₂₄В_{81,84}Г₅₄ — формула аркозового песчаника.

Формальное сопоставление записей этих двух горных пород указывает на определенное сходство между ними по минеральному составу и резкое различие в структуре (С₃ и С₂₄). Некоторые минералы одинаковые (81 — кварц, 54 — слюда), другие имеют близкие номера и относятся к одной группе (84 — полевой шпат, 87 — олигоклаз, 92 — КППШ).

С помощью семантических множителей составлены «эталонные» описания для 75 наименований горных пород. Каждому эталону присвоен условный числовой код, который используется для представления горных пород, состоящих из обломков других пород типа туфобрекчий и конгломератов, и в массивах фактической информации. Список эталонов является формализованным представлением названий горных пород, с его помощью осуществляется автоматическое установление названия любой горной породы, представленной семантическим кодом по ее свойствам.

Представление горных пород на ИПЯ указанным способом [78] обеспечивает:

- сопоставимость горных пород по их свойствам;
- возможность записать сведения о горной породе до установления ее названия;
- автоматизировать процесс установления названий;
- записывать и сопоставлять сведения о горных породах на разных уровнях детальности и ряд других преимуществ.

Рудные формации

Рудные формации относятся к категории эмпирических обобщений, позволяющих одним словом сообщить целый набор элементарных характеристик. В языке АФИПС эти понятия можно применять только после предварительной формализации, так как иначе будут возникать неоднозначные ситуации при записи сведений о конкретных объектах и при поиске данных.

В известных в настоящее время классификациях рудных формаций их общее число и отдельные свойства не совпадают. Более того, нет единой точки зрения на место рудных формаций среди других понятий, выражающих уровни организации вещества [26, 43]. Причина подобной неопределенности указана в Геологическом словаре [19, с. 384]: «Большое количество критериев, предлагаемых для определения понятия рудная формация, и различная оценка значения отдельных критериев затрудняет отнесение конкретных месторождений к той или иной рудной формации и приводят к выделению неодинакового количества рудных формаций разными исследователями».

Для формализации рудных формаций прежде всего нужно иметь достаточно четкое определение этого понятия. Наиболее распространенное сейчас в геологической литературе [43] определение рудной формации как группы месторождений со сходными по составу минеральными ассоциациями и близкими геологическими условиями образования в данном случае не подходит, так как не удовлетворяет необходимым требованиям по экспликации содержания. Это связано с тем, что в данном определении участвуют признаки разных уровней обобщения эмпирический факт — «минеральные ассоциации», которые можно установить достаточно однозначно, и теоретическое положение «условия образования», которые однозначно установить нельзя, а само понятие в определении не раскрыто и условия не перечисляются. С формальной точки зрения утверждение об одинаковых условиях образования должно быть скорее следствием из сходства в наблюдаемых признаках, но не собственно определяющим признаком, поскольку его наблюдать нельзя. Если же выбросить из приведенного определения вторую часть, то получится, что рудную формацию характеризует в основном минеральный состав, что может быть и справедливо, но с этим, видимо, не согласятся многие геологи.

Имеются также определения конкретных рудных формаций,

например такие, как определение колчеданной формации, данное С. Н. Ивановым [35]. В соответствии с этим определением колчеданная формация описывается целым рядом основных и менее характерных признаков. В числе основных — локализация оруденения в вулканогенных породах, согласное залегание рудных тел, серицитизация боковых пород и т. д. Все признаки, указанные в характеристике, являются вполне однозначными и устанавливаемыми эмпирически. Набор подобных признаков в совокупности позволил бы однозначно определить рудную формацию. Но данное определение содержит много избыточных признаков, а нужны только достаточные и необходимые. Кроме того, для большинства рудных формаций таких однозначных определений нет. В связи с этим выделение существенных признаков рудных формаций можно попытаться сделать на основе анализа их номенклатур.

В названиях рудных формаций встречается главным образом три типа характеристик: 1) полезный компонент (например, марганцевая, ртутная, медно-молибденовая, свинцово-цинковая и др.); 2) минеральный состав (сульфидно-касситеритовая, апатит-магнетитовая, кварц-касситеритовая); 3) вмещающие породы (магнетитовая скарновая, вольфрамовая грейзеновая). Из двух характеристик, относящихся к вещественному составу, в названиях всегда участвует по крайней мере одна; вмещающие породы в названиях часто отсутствуют. Естественно, что данные признаки являются, если не существенными, то во всяком случае главнейшими, и их можно принять за основу при экспликации понятий рудных формаций.

Соответственно каждая рудная формация может быть представлена в виде набора семантических множителей, принадлежащих к понятиям трех категорий — элементов, минералов и горных пород. Кодовая формула типа рудной формации, вводимая в ИПЯ как эталон названия, имеет следующий вид:

$$PФ = \{\mathcal{E}_i\} \{M_j\} \{BP_k\},$$

где $\mathcal{E}_i \in \mathcal{E}$; \mathcal{E} — множество полезных компонентов; $M_j \in M$; M — множество минералов, слагающих рудную массу; $BP_k \in GP$; GP — множество классов горных пород, которые являются вмещающими для рудных формаций.

Возможно также введение количественных определителей для компонентов и минералов, если окажется, что разные рудные формации не будут отличаться самым набором семантических множителей.

Каждый эталон рудной формации, вводимой в ИПЯ, получает свой условный код, которым можно пользоваться наряду с семантическим кодом. Присутствие в составе тезауруса системы эталонных описаний рудных формаций даст возможность автоматически изучать разные случаи отклонения в геологической характеристике конкретных объектов от определенного типа

рудной формации и сравнивать объекты между собой. Использование семантического кода по элементному и минеральному составу рудных формаций позволяет, кроме того, сопоставлять вещественный состав объектов на трех уровнях — элементов, минералов и рудных формаций.

Этот вариант представления рудных формаций в ИПЯ является вариантом *ad hoc* и не претендует на полную и истинную экспликацию этих понятий. Формальное сопоставление терминологии по рудным формациям свидетельствует о том, что скорее всего они принадлежат к тому же уровню организации вещества, что и горные породы, и поэтому по аналогии с ними должны определяться минеральным составом и структурой. Но так как понятий «структура рудных формаций» не существует, то полной аналогии с горными породами в настоящее время в отношении семантического кодирования рудных формаций не получается.

Геологические формации

Геологические формации, как и рудные, являются характеристикой, используемой для описания территориальных геологических объектов. Введение названий геологических формаций в ИПЯ имеет смысл только при условии их формализации. Геологические формации относятся к более высокому уровню организации вещества, чем горные породы [49], которые являются для них элементарными составляющими. Если при анализе геологических формаций для формализации и введения в ИПЯ опираться на определение их как парагенезисов горных пород, данное Н. С. Шатским [90], то каждый тип геологической формации можно представить как конъюнкцию семантических множителей, в качестве которых выступают отдельные горные породы. Здесь, как и при описании рудных формаций, по аналогии с другими вещественными категориями нужно было бы иметь понятие «структуры геологической формации», представляющее собой понятие более высокого порядка, чем понятие «структура горной породы». Однако такие понятия в большинстве случаев отсутствуют.

Из названий геологических формаций можно выявить еще два их признака — фациальные условия образования (например, покровно-ледниковая, красноцветная континентальная, лагунная соленосная, молассовая и т. д.) и условия залегания (платформенная, геосинклинальная). Возможно, что следует ввести эти признаки в качестве семантических множителей по крайней мере для некоторых формаций.

Таким образом, для рассмотренных выше категорий понятий вещественного состава наилучшим вариантом представления на ИПЯ является способ, позволяющий учитывать естественную иерархию этих объектов с помощью принципов семантического кодирования.

Процессы

Геологические процессы — это понятия, отображающие способы и условия образования различных геологических тел, в том числе минералов, горных пород, тектонических структур.

Имеется семь основных аспектов, с помощью которых раскрывается содержание понятий подавляющего большинства геологических процессов. Эти аспекты можно положить в основу набора семантических множителей, а также в качестве классификационных критериев классификаций процессов при дескрипторном представлении этих понятий.

Эти аспекты следующие:

- A* — вид процесса. $A = \{a_1 — \text{образование, } a_2 — \text{изменение, } a_3 — \text{разрушение}\}$.
B — тип процесса. $B = \{b_1 — \text{механическое (перемещение), } b_2 — \text{физическое (изменение структуры), } b_3 — \text{химическое (изменение состава), } b_4 — \text{физико-химическое (изменение структуры и состава)}\}$.
C — исходный объект. $C = \{\text{множества: } C_1 — \text{элементов, } C_2 — \text{минералов, } C_3 — \text{кристаллов, } C_4 — \text{горных пород, } C_5 — \text{веществ, } C_6 — \text{тектонических форм, } C_7 — \text{геологических тел}\}$. Например, $C_1 = \{\text{песчаник, гранит, ...}\}$, $C_3 = \{\text{магма, ...}\}$, $C_6 = \{\text{складка, сброс, ...}\}$, $C_7 = \{\text{пласт, жила...}\}$.
D — конечный объект. $D = C$ (те же объекты).
E — действующая сила. $E = \{T, P, \text{ гидротермы, магма, гравитация, вода}\}$. Например, $T = \{\text{высокая } t, \text{ средняя } t, \text{ низкая } t\}$, $P = \{\text{одностороннее, гидростатическое}\}$.
F — локализация в пространстве. $F = \{\text{приурочено к контакту, площадное распространение, точечное, поверхностное, глубинное, ...}\}$.
G — локализация во времени. $G = \{\text{одноактный, многоактный, моментальный, длительный}\}$.

Полная формула для описания любого понятия, относящегося к геологическому процессу

$$\text{Пр} = a_i^h b_j^k c_l^m d_n^p e_q f_r g_s.$$

Не для всех процессов используется полный комплект аспектов. Так, понятия «экзогенные» и «эндогенные» описываются только терминами из множества *F* (поверхностные, глубинные). Для других понятий достаточно и необходимо использование нескольких или всех групп семантических множителей.

Примеры.

Нефелинизация (превращение пироксенов в нефелин)*: $a_1 b_4 c_{\text{пир}}^2 d_{\text{неф}}^2$.

Пирометасоматизм (замещение одних минералов другими при выс. *t*)*: $a_2 b_4 c^2 d^2 t_{\text{выс}}$.

Палингенез (ультраметаморфический процесс образования магмы путем переплавления горных пород на глубине)*: $a_1 b_4 c^4 d_{\text{магма}}^5 t_{\text{выс}} f_{\text{глуб}} g_{\text{длит}}$.

Приведенная форма описания процессов нужна для тезауруса (ИПС) системы, а в массиве фактов они, как и другие сложные понятия, могут заменяться условным цифровым кодом.

Формы залегания и другие свойства геологических тел

Терминологию этой категории понятий целесообразно в ряде

* Толкование терминов дается по геологическому словарю [18, 19].

случаев разделить по их отношению к уровням организации вещества. Например, термины «кристалл», «конкреция», «друза», «гнездо» относятся к уровню описания минералов. Термины «покров», «пласт», «шток», «дайка» относятся к уровню горных пород. Для описания рудных тел терминология берется из обеих групп — «гнездо», «пласт», «шток», «линза» и т. д.

Дальнейшая экспликация смысла этой группы терминов достигается двумя способами: прямым закреплением некоторого смысла за термином или путем указания на его отношение к другим терминам той же категории. В число существенных признаков, характеризующих различные геологические тела, входят форма, размер, характер контактов, структура, тип слагающего вещества, генезис и др. Эти признаки учитываются в ИПЯ только в том случае, если по ним будут сравниваться различные геологические образования. Так, например, в определении даек [18] включен признак — «прямые контакты». Однако, по-видимому, этот признак не будет в числе поисковых для большинства ИПС, где фиксируются и описываются дайки. Поэтому этот аспект понятия «дайка» можно в ИПЯ не эксплицировать. Для многих терминов, относящихся к геологическим телам, достаточно ввести разграничивающие признаки или указание на отношения типа совместности и несовместности.

Например, целесообразно термин «слой» отнести к группе терминов стратиграфических подразделений и не вводить в число синонимов к термину «пласт»; между терминами «покров» и «пласт» провести разграничение по типу горных пород, к которым они относятся. «Покров» применим к эффузивам, а «пласт» — либо к осадочным, либо к метаморфическим породам. На этом основании между последними терминами можно установить отношение несовместности, а с точки зрения формы они совместны. Термины «покров» и «поток» относятся к горным породам и для них различающим признаком является форма. Чтобы эксплицировать форму, нужно ввести количественные соотношения для размеров этих тел, но так как это не всегда просто, то остается только указать в ИПЯ, что эти термины разные, но находятся в отношении совместности. Таким образом, установление типа отношений между этими терминами обусловлено в большой степени аспектами возможного рассмотрения соответствующих понятий в ИПС. Что же касается самой формы тел, то чаще всего средствами обычного геологического языка ее эксплицировать нельзя. Для этого требуется введение некоторых специальных понятий, которые обычно в геологических описаниях не встречаются (см. работы Ю. А. Воронина и Э. А. Еганова [16, 17]).

Остальные свойства геологических тел представляются на ИПЯ с помощью нормализованного перечня наименований с указанием некоторых отношений между ними, т. е. дескрипторным способом. При этом во многих случаях полная экспликация

свойств не требуется, так как элементарные понятия, характеризующие эти свойства, чаще всего выходят за пределы собственно геологических понятий и относятся либо к физике, либо к химии.

Стратиграфические подразделения

Представление терминов этой категории в ИПЯ опирается на стратиграфическую классификацию — Международную стратиграфическую шкалу и региональные стратиграфические шкалы. Для введения этих терминов в ИПЯ нужен полный перечень подразделений с указанием отношений между ними типа часть — целое и отношений раньше — позже, отражающих общую последовательность напластований. Все другие виды характеристик относятся к синтагматике языка.

Тектонические структуры

Тектонические структуры могут быть объектами специального рассмотрения в ФИПС, т. е. объектами, к которым относится информация, но главным образом они используются при описании других геологических объектов — территорий, рудных тел, месторождений с точки зрения их структурного положения и строения.

К существенным признакам тектонических структур относятся: размер, морфология, генезис, пространственное положение в структуре более высокого порядка, особенности элементов структуры; к элементам структур — замок, осевая поверхность, шарнир, крылья и др. Полная экспликация понятий тектонических структур только средствами обычного геологического языка невозможна. Предложенный Ю. А. Ворониным и др. [15, 16] способ формального описания структур не позволяет адекватно перевести язык существующих геологических описаний на этот формальный язык в связи с тем, что в последнем используются некоторые новые понятия. Положение с тектонической терминологией, очевидно, наиболее сложное в геологии [82, 48]. В работе Л. И. Красного [48] указывается, что расхождения в терминологии у разных исследователей приводит к «несбивке» листов геологической карты при полистном ее составлении. Учитывая это, для представления тектонических понятий на ИПЯ целесообразно иметь не только перечень вводимых в ИПС терминов и их синонимов, но и толковые определения для каждого термина, независимо от того, будут эксплицированы существенные признаки или нет, а также типы отношений между терминами. Для терминов этой категории вводятся следующие типы отношений: часть — целое, род — вид, совместность — несовместность. Без такой, хотя и очень ограниченной, экспликации содержания пользоваться тектоническими терминами в ИПЯ невозможно.

Перечисленные выше группы понятий и соответствующих им терминов составляют основу лексики и парадигматики ИПЯ для фактографических систем. Кроме них в состав ИПЯ входят также некоторые термины из общей лексики, для которых представление в геологической ФИПС является дескрипторным.

Синтагматика ИПЯ, способы построения высказываний

В языке фактографической ИПС можно использовать все известные типы синтагматических средств. При этом для полной передачи смысла высказываний, как правило, нужны более точные средства, чем только указатели роли или связи. Из опыта построения информационно-поисковых языков в некоторых областях знаний известно, что содержание текста номинативных предложений типа заголовков можно передать достаточно однозначно с помощью модели языка объектно-предикатного типа, т. е. состоящего из объектов и отношений [47, 76]. В геологии с помощью подобного языка можно записывать не только номинативные сочетания, но и в ряде случаев конкретную геологическую информацию.

Рассмотрим основные принципы построения языка объектно-предикатного типа, позволяющего записывать геологический текст, содержащий фактические сведения об объектах.

В геологических текстах такого типа при построении предложений (высказываний) используются в основном термины трех групп.

1. Термины, обозначающие геологические объекты и их свойства (рассмотрены в предыдущем разделе).

2. Слова общей лексики, указывающие на геологическую ситуацию, в которой находятся эти объекты, а также на их взаимоотношения между собой.

3. Специфические геологические термины, характеризующие геологический аспект рассмотрения объектов.

К последним относятся, например, такие термины: «вмещающие породы», «ксенолиты», «главные минералы рудного тела», «аксессуарные минералы», «лежачий бок» и т. д. Эти термины не обозначают самостоятельные геологические объекты, как термины первой группы, а лишь указывают на нахождение этих объектов в определенной ситуации или в определенном качестве.

Слова второй и третьей групп можно представить в информационно-поисковом языке как отношения между объектами (терминами первой группы).

Важнейшие типы отношений между геологическими объектами условно можно отнести к четырем группам: пространственные; временные; вещественные; отношения действия, состояния и пр. Им соответствуют синтагматические отношения между терминами языка. Примеры этих отношений приведены в табл. 6.

Примеры отношений между терминами, обозначающими геологические объекты

Примеры отношений Между объектами	Условный код	Обратные отношения	Примеры и примечания
Пространственные отношения			
Находиться в	S_1^*	Находиться за (за пределами)	Объект* находится в пределах n -й структуры
Находиться выше	S_2	Находиться ниже	Песчаник залегает под угольным пластом
Иметь пространственное отношение (приурочены)	S_3	Пространственная приуроченность отсутствует	Жилы приурочены к секущим трещинам
	S_4		Трехместное отношение
Находиться между		—	
Находиться в верхней части	S_5	Находиться в нижней части	
Находиться на глубине	S_6	Находиться на поверхности	
Находиться в центре	S_7	Находиться на периферии	
Уточнение места	S_8	—	Минерал находится в висячем боку рудного тела

Временные отношения

Иметь отношение (временное)	T_1	Не иметь временной связи	Месторождение приурочено к девонским отложениям (двухместное отношение)
Быть одновременно Быть раньше Быть в геологическое время (иметь геологический возраст)	T_2	Быть разновременно	Отличается от T_1 . Месторождение образовалось в девоне (одноместное отношение)
	T_3	Быть позже	
	T_4	—	
Иметь абсолютный возраст	T_5	—	

Вещественные отношения

Иметь в своем составе (состоит из)	M_1	Является составной частью	Кварцевая жила (жила из кварца)
------------------------------------	-------	---------------------------	---------------------------------

Отношения действия, состояния и прочие

Иметь некоторое свойство	E_1^i	Быть свойством	Это большая группа отношений, в которую входят: «быть вмещающей породой», «иметь залегание» и т. п.
--------------------------	---------	----------------	---

Примеры отношений между объектами	Условный код	Обратные отношения	Примеры и примечания
Быть объектом (рассмотрения или действия)	E_2	Быть субъектом	Это отношение не затрагивает вещественный состав Много, мало, несколько, 10, 6—9 и т. д.
Быть причиной	E_3	Быть следствием	
Иметь условие	E_4	Быть условием	
Включает	E_5	Принадлежит (относится к)	
Быть больше	E_6	Быть меньше	
Быть в количестве	E_7		
Иметь сходство	E_8	Не иметь сходства	
Уточнение отношения (любого из этой группы)	E_9	—	
Переходное состояние (переход к)	E_{10}	—	
Иметь генетическое отношение	E_{11}	Не иметь генетического отношения	

* Обратные отношения кодируются теми же символами, но со знаком отрицания.

Рассмотрим на примере, как с помощью названных типов отношений можно передавать основное содержание геологического текста.

Возьмем фрагмент текста, взятый из работы А. В. Коплуса «О натровом метасоматозе вмещающих пород и роли альбитизирующих растворов в формировании хрусталеносной минерализации Алданского щита» («Изв. вузов. Сер. Геология и разведка», 1974, № 6, с. 43.).

«В западной части Алданского щита широко распространены представляющие практический интерес проявления горного хрусталя, связанные с кварцевыми жилами гидротермально-метаморфического типа. Все они приурочены к кварцитам верхнеалданской свиты архея, породы которого метаморфизованы в условиях гранулитовой и амфиболитовой фаций».

Используя указанные термины, обозначающие объекты и

Геологические объекты и их свойства (термины первой группы)		Отношения	
Проявления	a_1	Находиться в	S_1
Горный хрусталь	a_2	Иметь пространственное отношение	S_3
Алданский щит	a_3		
Жилы	a_4	Уточнение места	S_8
Кварц	a_5	Иметь геологический возраст	T_4
Гидротермальный	a_6	Иметь в составе	M_1
Метаморфизм	a_7	Иметь свойства	E_1
Кварциты	a_8	Быть объектом	E_2
Верхнеалданская свита	a_9	Быть результатом процесса	\bar{E}_3

Геологические объекты и их свойства (термины первой группы)		Отношения	
Архей	a_{10}	Принадлежать	\bar{E}_5
Гранулитовая фация	a_{11}	Быть в количестве	E_7
Амфиболитовая фация	a_{12}	Иметь генетическое отношение	E_{11}
Запад	a_{13}		
Практический интерес (перспективность)	a_{14}		
Широко распространены (много)	a_{15}		

типы отношений, смысл приведенного выше текста, разбив его на части в соответствии с объектом описания (подлежащим), можно записать следующим образом.

$$1. E_2(a_1M_1a_2)E_7a_{15}E_{14}S_{14}S_8a_{13}S_3(a_4M_1a_5)S_3a_8.$$

Здесь записано: проявления горного хрусталя в большом количестве, перспективны, находятся в Алданском щите, его западной части, связаны с кварцевыми жилами, имеют пространственную связь с кварцитами.

$$2. E_2(a_4M_1a_5)\bar{E}_3(a_7E_1a_6).$$

Записано: кварцевые жилы являются следствием (процесса) гидротермального метаморфизма.

$$3. E_2a_8\bar{E}_5(a_9T_4a_{10}).$$

Записано: кварциты принадлежат к верхнеалданской свите архея.

$$4. E_2(a_9T_4a_{10})\bar{E}_3(a_7E_1a_{11}\&a_{12}).$$

Записано: верхнеалданская свита подверглась метаморфизму в гранулитовой и амфиболовой фациях.

Правила расшифровки фраз следующие. На первом месте стоит одноместное отношение E_2 , указывающее объект рассмотрения (подлежащее фразы). Далее читается текст таким образом, что между объектом рассмотрения и каждым из последующих объектов устанавливается отношение, стоящее перед этим объектом. Если объект имеет определение (одноместное отношение), тогда он вместе с определением заключается в скобки. Отношение уточнения не является самостоятельным, оно уточняет отношение, стоящее перед ним.

С помощью указателей связи можно ввести и другие правила записи и чтения фраз, например разбить весь текст на синтагмы [117]. Это вполне возможно, так как большинство отношений являются двухместными, а одноместное отношение можно преобразовать в двухместное. Но в последнем случае будут повторяться записи одного и того же объекта в нескольких синтагмах, а, кроме того, приведенная выше запись, хотя и состоит из двухместных отношений, но в целом связывает между собой все термины фразы.

В некоторых системах сами отношения могут находиться между собой в смысловой связи; среди них выделяются простые и производные от них отношения [76].

Язык, построенный подобным образом, обладает достаточной однозначностью и семантической силой и позволяет получать логические выводы о связях между объектами и их свойствами. Однако недостатком такого языка является сложность перевода с ЕЯ на ИПЯ. Перевод усложняется, если в тексте встречается информация разного уровня, относящаяся к разным объектам, а также в тех случаях, когда в тексте имеется большое число рассуждений. Перевод на ИПЯ намного легче, если текст на ЕЯ стандартизован и относится к объектам одной категории. Если структура всех фраз на ИПЯ одинакова или повторяется, то мы имеем грамматику стандартных фраз. При этом символы, обозначающие отношения в массиве фактов, каждый раз не записываются, а только подразумеваются. Запись грамматической структуры фраз хранится в памяти ЭВМ и учитывается при формулировании запросов и ответов.

В настоящее время структура записи фактографической информации в виде стандартных фраз является наиболее приемлемой. Пример ИПЯ, в котором реализована грамматика типа стандартных фраз, рассмотрен в следующем разделе.

Пример ИПЯ с синтагматическими отношениями типа «стандартных фраз»

ИПЯ разработан для описания рудных месторождений в автоматизированной АФИПС [31, 81]. Массив информации состоит из записей. Каждая запись содержит сведения об одном месторождении. Синтаксическая структура записи представляет собой сложную стандартную фразу (мегафразу), состоящую из стандартных фраз различных уровней. Число мест (переменных) в стандартных фразах каждого уровня фиксировано — это число аспектов описания объекта. Областью значений переменных являются либо стандартные фразы низшего уровня, либо единичные термины ИПЯ, либо термины вместе с указателями роли. Термины представлены в двух вариантах: дескрипторным способом и на языке семантических множителей. На множестве единичных терминов заданы отношения эквивалентности, совместности, несовместности, частичного сходства и подчинения.

Мегафраза в целом может считаться стандартной фразой третьего уровня, она включает 9 стандартных фраз второго уровня Φ_1 — Φ_9 . Объектом описания в Φ_1 , Φ_2 и Φ_3 является рудное тело. Объектом описания в Φ_5 являются вмещающие породы месторождения. Объектом в Φ_8 являются горные породы вблизи месторождения. Объектом в Φ_4 , Φ_7 и Φ_9 является месторождение в целом.

Стандартные фразы второго уровня состоят из стандартных фраз первого уровня — (T'_i) , единичных терминов и терминов с указателями роли — T'_i . При этом каждое место может

быть занято не одним термином или стандартной фразой, а несколькими. Элементы информации в стандартных фразах разделяются между собой скобками и знаками пунктуации*

$$\Phi_1 = [T_1^1; T_1^2; T_1^3; T_1^4; T_1^5];$$

$$\Phi_2 = [(T_2^1); T_2^2; T_2^3; T_2^4; T_2^5];$$

$$\Phi_3 = [(T_3^1); (T_3^2); T_3^3; T_3^4; T_3^5; T_3^6; T_3^7; T_3^8];$$

n раз n раз

$$\Phi_4 = [T_4^1; T_4^2; T_4^3; (T_4^4); (T_4^5); (T_4^6)];$$

$$\Phi_5 = [T_5^1; T_5^2; T_5^3; T_5^4; T_5^5; (T_5^6); T_5^7; T_5^8]; \text{ n раз.}$$

$$\Phi_6 = [T_6^1; (T_6^2); T_6^3; T_6^4; T_6^5; T_6^6; T_6^7; T_6^8; T_6^9];$$

$$\Phi_7 = [T_7^1; T_7^2; T_7^3; T_7^4; T_7^5; T_7^6; T_7^7; T_7^8; T_7^9];$$

$$\Phi_8 = [(T_8^1); (T_8^2)];$$

n раз n раз

$$\Phi_9 = [T_9^1; T_9^2; T_9^3; T_9^4; T_9^5; T_9^6; T_9^7; T_9^8; T_9^9; T_9^{10}; T_9^{11}; T_9^{12}; T_9^{13}];$$

Данный способ выражения синтагматических связей позволяет передать следующие виды отношений между терминами: иметь форму — T_1^1 , иметь размер — T_1^2 , находиться в пределах — T_1^9 и более сложные, характеризующие вещественный состав рудных тел, распределение в них компонентов и др.

Например, (T_2^1) — означает множество высказываний, характеризующих распределение компонентов в рудном теле. Эти высказывания образуются из терминов следующих категорий: вид распределения = {равномерное, неравномерное, вертикальная зональность, ... и т. п.};

вид изменения содержания компонентов = {увеличение, уменьшение};

направление изменения = {на глубину, к поверхности, к центру, ... и т. п.};

вид компонента = {множество элементов и минералов}.

Высказывание: «увеличение с глубиной халькопирита» записывается как $9a \text{ CuFeS}_2$.

Стандартная фраза (T_3^1) раскрывает многоместное отношение типа: x_i есть главный рудный минерал, находящийся в количестве $y_i\%$, содержит примеси элементов z_i в количестве k_i , z_j в количестве k_j , ..., имеет габитус l_i и текстурно-структурные особенности типа m_i . В этой фразе областью значений x_i является множество промышленных рудных минералов [32], областью значений y_i — множество целых чисел до 100, областью значений z_i — множество элементов-примесей, областью значений k_i — множество действительных чисел, больших нуля, областью значений l_i — множество терминов, обозначающих

* В квадратные скобки заключены стандартные фразы второго уровня, в круглые — стандартные фразы первого уровня.

габитусные формы минералов, областью значений m_i — множество терминов, обозначающих текстурно-структурные особенности минералов.

Приведем общий перечень сведений, который может быть записан о месторождении в данном ИПЯ.

Ф₁ (форма и размещение рудного тела в пространстве) — морфология, размеры, элементы залегания, способ залегания, характер контактов.

Ф₂ (особенности строения рудного тела) — размещение компонентов, ксенолиты вмещающих пород, текстура рудных агрегатов, структура рудных минералов, метаморфизм руды.

Ф₃ (вещественный состав руды) — минералы главные рудные, их количество в рудной массе, элементы-примеси, количество примеси, габитусные формы, текстурно-структурные особенности; минералы главные жильные, их количество, элементы-примеси; минералы второстепенные рудные; минералы рудные редкие; минералы жильные прочие; минералы вторичные; полезные элементы в руде и их содержание; геохимическая характеристика руды, сравнения минералов.

Ф₄ (общее строение месторождения) — количество рудных тел, протяженность оруденения по вертикали и в плане, изменение компонентного состава в разных рудных телах, зона окисления, ее мощность, минеральный состав, полезные компоненты; те же сведения о зоне цементации.

Ф₅ (характеристика вмещающих пород) — название; состав; структура; залегание; метаморфизм; форма залегания; ореолы рассеяния рудных компонентов, состав, размер и форма ореола; возраст абсолютный и геологический.

Ф₆ (околорудные изменения вмещающих пород) — интенсивность изменения, состав зон, мощность, сравнение всячего и лежащего боков, изменение физико-механических свойств, зональность, пространственное соотношение оруденения и измененной зоны, баланс элементов.

Ф₇ (геолого-структурное положение месторождения) — контроль оруденения, возраст оруденения абсолютный и геологический, структурная позиция месторождения, структурный контроль рудных тел, региональное положение месторождения, характеристика рудовмещающего комплекса, его геологический возраст, формация горных пород.

Ф₈ (характеристика горных пород месторождения) — горная порода по составу и структуре или по названию, ее отношение к вмещающей породе, характер контактов с вмещающей породой, форма залегания, метаморфизм. Те же сведения о дайковых породах.

Ф₉ (общие и экономические сведения о месторождении) — географическое положение, координаты, название, начало и окончание работ, степень изученности и разведанности, способ разработки, масштаб, запасы, оценка качества руды по содержанию, глубина залегания и глубина разведки, пострудные смещения.

Глава 4

МЕТОДЫ СРАВНЕНИЯ ИНФОРМАЦИИ В ФАКТОГРАФИЧЕСКОЙ ИПС

Общие замечания

Автоматический поиск заключается в том, что считывающее устройство ЭВМ просматривает элементы информации, находящиеся в ее памяти, и те из них, которые отвечают условиям запроса, фиксируются и поступают на выход. В процессе сравнения элементов информации из массива с указанными в запросе требованиями в общем случае может происходить сортировка

информации на три группы: информацию, соответствующую запросу; информацию, не соответствующую запросу; информацию, соответствующую запросу неполностью.

Степень соответствия устанавливается по формальным признакам или с использованием специально разработанных правил, учитывающих как априорные, так и ситуативные отношения между сравниваемыми элементами информации.

В содержательном аспекте установление степени соответствия означает, например, следующее.

Предположим, что по некоторому запросу требуется выбрать объекты, залегающие в гранитах. При автоматическом просмотре описаний объектов, содержащихся в памяти ИПС, по всей вероятности, окажется, что наряду с объектами, залегающими в гранитах и явно не в гранитах, есть объекты, залегающие в гранодиоритах, в гранит-порфирах, граносиенитах, т. е. в породах, близких к запрашиваемым. Система должна решить, какие из них соответствуют запросу и подлежат выдаче, а какие нет.

При поиске по совокупности свойств чаще всего окажется, что часть характеристик будет совпадать с условиями запроса, а часть — совпадать лишь частично. И в этом случае формальные правила предписывают, когда и какую информацию нужно выдавать. Вопрос о соответствии между элементами информации не всегда решается достаточно просто.

Известно, что степень сходства между какими-либо предметами, явлениями или свойствами всегда рассматривается с позиций определенного уровня, т. е. с определенной точностью. Информация, которая может считаться равнозначной при сравнении на одном уровне, при сравнении на другом уровне будет считаться различной. Уровень, на котором ведется сравнение, обусловлен прежде всего априорными знаниями относительно этих предметов или свойств.

В ИПС имеем то же самое. Максимально возможная детальность рассмотрения обусловлена семантическими возможностями ИПЯ. Если имеется несколько уровней представления информации, каждый из которых раскрывает содержание информации с разной степенью точности, то должно также быть и несколько уровней, на которых устанавливается соответствие.

В тех случаях, когда отношения между терминами ИПЯ не заданы, при сравнении двух терминов может быть только два состояния — совпадение и несовпадение. При наличии отношений между терминами число возможных состояний больше двух. При сравнении числовой информации устанавливаются отношения равенства, неравенства (больше, меньше). При поиске по нескольким терминам, даже в случае отсутствия отношений между ними, будут возникать более сложные ситуации, чем просто совпадение и несовпадение.

Помимо семантических особенностей ИПЯ на детальность

сравнения влияет фактическая потребность, определяемая по характеру запроса. В ряде случаев фактическая потребность в определенной точности рассмотрения информации может быть ниже максимально возможной детальности, обусловленной составом информации и уровнем языка.

Учитывая разнообразный характер запросов в фактографических ИПС не всегда нужно рассматривать сходство между понятиями с максимально возможной детальностью. Во-первых, слишком большая детальность может затушевать нужное сходство, а, во-вторых, более детальное рассмотрение требует увеличения количества операций при сравнении, что увеличивает время поиска. В каждом конкретном случае существует, по-видимому, оптимальная детальность сравнения. Поэтому в фактографических системах при разработке правил установления смыслового соответствия между единицами информации и объектами целесообразно предусмотреть возможность сравнения с разной степенью точности, чтобы использовать это при решении различных задач.

При поиске и сравнении между собой объектов по комплексу свойств в ФИПС можно применять как собственно информационные методы, основанные на использовании той информации, которую дает ИПЯ, так и методы математические, в которых содержательная сторона каждого свойства не рассматривается.

Установление степени сходства между двумя элементами информации в АФИПС

Рассмотрим правила установления степени сходства для двух основных способов представления понятий в ФИПС с указанием их смысловых связей — дескрипторного и по семантическим множителям [79].

Для понятий, представленных дескрипторным способом, максимальное число типов соответствия равно трем: полное соответствие, частичное соответствие и взаимоисключение. Критериями их установления являются виды отношений между терминами, зафиксированные в ИПЯ. Критерием полного соответствия является совпадение двух терминов и вариант, при котором термин запроса является более общим, чем термин массива данных. Критерием частичного соответствия является либо наличие в запросе термина более частного по отношению к сравниваемому с ним термину массива фактов, либо, если сравниваются два термина из массива фактов, наличие между ними отношения подчинения одного другому. Критерием взаимоисключения двух терминов является наличие между ними специально зафиксированного в ИПЯ отношения несовместности. Функция соответствия «взаимоисключение» состоит в следующем. Если в процессе поиска по некоторому свойству (термину) установлено для некоторого объекта отношение

взаимосключения, то этого свойства ожидать у данного объекта нельзя. Этот вариант поиска служит первой проверкой при прогнозировании свойств.

Степень сходства при сравнении терминов, представленных семантическим кодом, устанавливается при автоматическом сравнении на основе заранее составленных правил соответствия, учитывающих совпадение количества и вида семантических множителей.

Способ построения таких правил рассмотрим на примере горных пород. В соответствии с представлением горных пород по семантическим множителям, описанном выше, для них использовано десять стандартных типов сходства, фиксирующих разную степень близости сравниваемых понятий [79].

Тип I. Абсолютное сходство. Этот тип устанавливается при совпадении порядка расположения в формуле и вида всех семантических множителей горных пород.

Тип II. Полное сходство по структуре. Устанавливается при совпадении множителей, обозначающих структуры горных пород.

Тип III. Неполное сходство по структуре. Критерием этого типа сходства является наличие между терминами, обозначающими структуры, отношений подчинения и соподчинения. Этот тип сходства, как и предыдущий, свидетельствует о принадлежности горных пород к одной генетической группе (эффузивных, осадочных, пирокластических и т. д.).

Тип IV. Полное сходство по минеральному составу. Устанавливается при совпадении множителей, обозначающих породообразующие минералы, а также количественных и качественных определителей для этих минералов.

Остальные шесть типов (с V по X) сходства устанавливаются в зависимости от числа совпавших минералов и видов их определителей. Они соответствуют неполному сходству по минеральному составу. При этом сходство типа V имеют горные породы, у которых одинаковы все основные породообразующие минералы, сходство типа X — горные породы, у которых имеется только по одному общему минералу из числа аксессуарных.

Правила присвоения типа сходства горным породам по минеральному составу зафиксированы в специальной таблице, имеющей вид, показанный в табл. 7. По этой таблице, зная результат сравнения, т. е. число одинаковых минералов и виды их определителей, алгоритмически устанавливается тип сходства между двумя горными породами.

Описание алгоритма сравнения семантических кодов двух горных пород

Сравнение кодовых формул горных пород может проводиться либо по полному алгоритму, либо по сокращенному. Полный алгоритм позволяет в итоге оценить сходство по всем

имеющимся характеристикам (семантическим множителям) и установить сходство с максимальной точностью. Сокращенный алгоритм позволяет установить сходство либо по одной характеристике, либо по двум-трем.

Таблица 7

Вид таблицы для установления типов сходства горных пород по минеральному составу

Виды первых определителей минерального состава	Типы сходства горных пород по минеральному составу						
	IV	V	VI	VII	VIII	IX	X
Варианты							
А А							
А Б	—	МА=МБ МВ=МВ	МА=МБ	МА=МВ	МВ=МБ(В) и М=М	МВ=МБ(В) или 2М=2М	М=М
А В							
А Е							
Б Б							
Б В							
Б Е							
В В							
В Е							
Е Е							

Примечание. В позициях указывается наличие необходимого количества совпавших минералов с указанием вида их определителей. В качестве примера заполнена только строчка для пары первых определителей А. Б. Здесь МА=МБ означает, что минерал, стоящий с определителем А 1 горной породы (ГП), совпадает с минералом, стоящим с определителем Б 2 ГП. МВ=МВ означает совпадение минералов без учета их определителей. Прочерк позиции в IV АБ означает, что ситуация невыполнима (так как полное сходство не могут иметь ГП, имеющие разные первые определители).

Полный алгоритм сравнения двух горных пород начинается с проверки на совпадение терминов, обозначающих структуры. В случае несовпадения последние проверяются на отношение подчинения и соподчинения между ними. После сравнения структурных индексов сравниваются индексы минерального состава. Для этого сначала устанавливаются виды определителей, стоящих в формулах перед первыми минералами. Всего может быть десять вариантов (см. табл. 7), и соответственно алгоритм подразделяется на десять ветвей для каждой пары первых определителей. Далее идет процесс сравнения каждого минерала одной горной породы с каждым минералом другой горной породы, сначала с учетом только основных определителей, а затем, если нужно, и всех остальных. Результат сравнения оценивается в процессе выполнения алгоритма несколько раз для того, чтобы проверить целесообразность дальнейшей процедуры на предмет возможного увеличения степени сходства при последующем сравнении или отсутствия такой возможности.

Особенностью данного алгоритма является то, что при сравнении учитывается не только совпадение и несовпадение определенных семантических множителей, но и случаи возможной взаимозаменяемости некоторых минералов в однотипных горных породах. Все случаи допустимых замен зафиксированы в списке; при его составлении принимались во внимание как отношения между терминами типа подчинение и соподчинение, так и функции минералов в определении типа горной породы.

Например, в магматических горных породах номер плагиоклаза указывает на тип горной породы в гораздо большей степени, чем темноцветный минерал. Так, гранит может быть двуслюдяной или содержать один вид слюды. В качестве темноцветного в нем может быть не слюда, а роговая обманка. Но нет гранита, содержащего андезин, не говоря уже о более основных плагиоклазах. Поэтому с учетом названных особенностей недопустимой при установлении типа сходства горных пород является замена плагиоклазов различных номеров, а замена терминов, обозначающих моноклинный пироксен и ромбический пироксен, позволяет снизить уровень сходства на единицу. Термины «ортоклаз» и «микроклин» считаются при установлении сходства эквивалентными. Для плагиоклазов разрешается замена только для соседних номеров, причем с понижением уровня сходства. Используется также правило, что если в одной из двух сравниваемых горных пород присутствует минерал, являющийся более общим понятием, чем минерал в другой горной породе, то при сравнении соответствующие термины эквивалентно заменяют друг друга [79].

Не рассматривая здесь общую схему алгоритма, укажем последовательность операций на конкретном примере.

Допустим, нужно установить тип сходства для двух горных пород (ГП)

$$1. \text{ ГП} = C_{7;36}B_{77;106}Г_{81;84},$$

$$2. \text{ ГП} = C_{36;11}B_{86;17;41;75;56;106}.$$

Для наглядности раскроем содержание индексов: C_7 — структура лампрофировая, C_{11} — структура долеритовая, C_{36} — структура метасоматическая; минералы: 77 — клинохлор, 75 — хлорит, 106 — кальцит, 105 — карбонат, 81 — кварц, 84 — полевой шпат, 86 — альбит, 17 — эпидот, 41 — актинолит, 56 — мусковит.

Название операции	Результат
1. Проверка структурных индексов на совпадение	Совпадает один из индексов, указывающий на вторичную структуру
2. Проверка индексов первичных структур нахождение в одной группе	Находятся в одной группе изверженных пород
3. Оценка результата	ГП имеют сходство по структуре типа III
4. Установление индексов первых определителей ГП	В и В

Название операции	Результат
5. Переход к соответствующей ветви алгоритма с этой парой индексов (вариант 8)	
6. Составление таблицы исходных данных вида	
$\frac{1 \text{ ГП } \text{МВ} (\text{М}')}{2 \text{ ГП } \text{МВ}}$	$\frac{1 \text{ ГП } 77 (75), 106 (105)}{2 \text{ ГП } 86, 17, 41, 75, 56, 106}$
где МВ — минералы, стоящие с определителем В; (М') — минералы, эквивалентно их заменяющие	где минералы, стоящие в скобках, берутся из списка эквивалентных замен, в котором каждому минералу сопоставлены его возможные замены
7. Операция сравнения каждого минерала верхней строчки (1 ГП) с каждым минералом из нижней строчки (2 ГП)	Совпадает два минерала — фиксируется состояние Φ_3 , не совпадает более двух минералов — фиксируется состояние Φ_6 (в алгоритме указаны все возможные состояния)
8. Общий итог сравнения	$\Phi_3 + \Phi_6$ — сходство типа VII (указано в алгоритме)
9. Проверка целесообразности дальнейшего сравнения <i>Пояснение:</i> в сравнении еще не участвовали минералы 1 ГП, стоящие с определителем Г	Нет (оценка по результату, полученному в описании 8)
10. Выдача ответа	<i>Пояснение:</i> так как не совпадает более двух минералов, стоящих с определителем В, то тип сходства не изменится Сходство по минеральному составу типа VII
11. Общий ответ	Сходство по структуре типа III, по минеральному составу типа VII

Рассмотренные особенности сравнения двух горных пород по семантическим множителям показывают не только принципы сравнения понятий, записанных семантическим кодом, но и некоторые признаки автоматического логического анализа информации, когда при сопоставлении объектов учитывается не только формальное наличие или отсутствие некоторых свойств, но и отношения между ними, их возможная взаимозаменяемость. Логический анализ в рассмотренном случае достигается средствами информационно-поискового языка.

Сравнение объектов информации по комплексу свойств

Сравнение объектов по совокупности свойств и нахождение между ними сходства может проводиться аналогично сравнению между собой двух сложных терминов, представленных набором семантических множителей.

Критерий сходства объектов по совокупности свойств определяется с учетом сходства по каждому свойству и числу совпавших свойств. При этом не всегда можно заранее установить стандартные уровни сходства и предусмотреть в запросе все возможные случаи комбинаций свойств, которые окажутся в массиве фактов. Как правило, получение гибких ответов достигается двумя путями: уточнением запроса и его коррекцией в

процессе диалога между машиной и человеком и коррекцией ответа — выдачей информации с разной степенью соответствия. Оба пути тесно взаимосвязаны. Целесообразно сложный запрос анализировать с геологической точки зрения и только после этого в поисковом предписании указывать условия сравнения и общий критерий выдачи.

Подход к решению подобных задач поиска по совокупности свойств с учетом смысловых связей между терминами рассмотрим на одном из примеров [80].

Предположим, требуется назвать объекты, отвечающие следующим четырем свойствам.

1. Околорудные измененные породы представлены кварц-серицитовыми сланцами.

2. В составе жильных минералов имеется кварц и барит.

3. Рудные тела залегают согласно с вмещающими породами.

4. Главными рудными минералами являются пирит, халькопирит, сфалерит.

Логический анализ запроса показывает, что по свойству 1 возможны такие варианты:

a_1 — полное совпадение с заданным термином,

b_1 — частичное совпадение с заданным термином в сторону увеличения или уменьшения числа составляющих минералов в горной породе.

Например, возможны объекты, где околорудно-измененные породы представлены не кварц-серицитовыми сланцами, как запрашивается, а кварц-хлоритовыми, кварцитовыми, серицитовыми, кварц-серицит-хлоритовыми и т. д. Остальные случаи, естественно, можно не учитывать (полное несовпадение, совпадение по структуре и т. п.).

По свойству 2 имеет смысл принимать во внимание только следующие варианты:

a_2 — полное совпадение,

b_2 — присутствие только одного из названных минералов.

По свойству 3 целесообразно учитывать такие варианты:

a_3 — совпадение терминов,

b_3 — отсутствие сведений об этом свойстве наряду с отсутствием свойства, несовместного с ним.

По свойству 4, очевидно, нужно учитывать следующие варианты:

a_4 — полное совпадение,

b_4 — совпадение по двум минералам, из которых один должен быть пиритом, или присутствие кроме указанных в запросе еще одного минерала дополнительно.

Рассмотрев каждое свойство в отдельности, можно сформулировать общий критерий сходства. Максимальное сходство (первый уровень) имеют объекты, обладающие следующим набором свойств:

$$A_1 = a_1 \& a_2 \& a_3 \& a_4.$$

Второй уровень сходства имеют объекты, отвечающие следующим условиям по набору свойств:

$$A_3 = (a_1 \& a_2 \& a_3 \& b_4) \vee (a_1 \& a_2 \& b_3 \& a_4) \vee \\ \vee (a_1 \& b_2 \& a_3 \& a_4) \vee (b_1 \& a_2 \& a_3 \& a_4).$$

К третьему уровню сходства можно отнести объекты, у которых имеются несовпадение по двум свойствам, и т. д.

Таким образом, как показано на примере, критерий выдачи может быть разным. Возможность проводить сравнение с разной степенью соответствия моделирует содержательный геологический анализ и позволяет проводить аналогии между объектами в пределах разных по масштабам территорий и среди разного числа самих объектов. Выдача объектов, отвечающих определенному уровню сходства, представляет собой классификацию этих объектов. Так как классификация, получаемая таким способом, может быть гибкой, то ее можно использовать для различных задач.

Однако к установлению степени сходства объектов по комплексу признаков можно подойти и с более формальных позиций без анализа самой информации и без учета априорной значимости каждого свойства для той или иной задачи.

Имеется большое число различных математических методов установления мер сходства, мер связи или мер подобия. Прежде всего, как указано в работе Р. К. Кроса и др. [47], это классическая задача математической статистики, которая используется для этой цели коэффициенты корреляции, факторный анализ и другие методы.

Наибольшее применение нашли методы установления сходства на основе сравнения количественных данных, особенно петрохимических [1, 71], но используются также методы сравнения по качественным признакам [44, 23]. Аксиометрический подход к установлению меры сходства описан у Ю. А. Воронина [14].

Подробное рассмотрение математических методов не входит в нашу задачу. Укажем только основные принципы количественного выражения сходства между двумя объектами. Таких основных способов Н. Бейли [6] выделяет три: с помощью коэффициента ассоциации, коэффициентов корреляции и показателей расстояния.

Простейшим коэффициентом ассоциации для двух объектов является относительное число совпавших признаков

$$S = \frac{m}{n},$$

где m — число совпавших признаков; n — общее число признаков. Более сложные коэффициенты ассоциации учитывают также число несовпадений, однако эти коэффициенты не все-

да надежнее сравнительно простых [6]. Примером более сложного коэффициента ассоциации является коэффициент S_{sm} — Сокэла и Миченера (Sokal, Michener).

$$S_{sm} = (p+n):(p+n+u),$$

где p — число совпавших характеристик; n — число несовпавших характеристик; u — число несопоставимых характеристик. Применение S_{sm} для установления степени сходства вольфрамовых месторождений по набору из 26 признаков описано в работе П. Коллера и Д. Мериама [100].

Коэффициент корреляции используется для количественного измерения степени связи между признаками непрерывного характера и в меньшей мере для дискретных признаков. Формула для вычисления коэффициента корреляции приводится во всех учебниках по математической статистике, поэтому приводить ее здесь не будем.

Показатели «расстояния» характеризуют степень расхождения в значениях признаков сравниваемых объектов в многомерном пространстве. Обобщенное выражение для «расстояния» между двумя точками (объектами) j и k в n -мерном пространстве имеет вид

$$\Delta_{jk}^2 = \sum_{i=1}^n (x_{ij} - x_{ik})^2,$$

где x_{ij} и x_{ik} — значения некоторого признака i для объектов j и k соответственно. Этот способ применяется для оценки сходства по количественным признакам.

Установление мер сходства между парами объектов позволяет затем провести группирование объектов в классы сходства. Для этого удобно использовать матрицу сходства. Она имеет вид прямоугольной матрицы $n \times n$, где n — число объектов. В каждой позиции матрицы ставится коэффициент сходства между двумя объектами, полученный любым из способов. Чтобы выявить объекты, имеющие максимальную степень сходства, изучается разброс значений коэффициента сходства и в пределах этих значений выделяется несколько интервалов (порогов). По ним и происходит объединение в группы сходства.

Таким образом, математические методы могут выполнять ту же функцию, что и информационные способы, — группирование объектов по типу сходства.

Глава 5

ОБЩИЕ СВЕДЕНИЯ О ЭВМ И МАТЕМАТИЧЕСКОМ ОБЕСПЕЧЕНИИ СИСТЕМ

Основные требования к техническим средствам со стороны АФИПС и виды ЭВМ, используемые в настоящее время

Выбор технических средств обусловлен параметрами фактографической информационно-поисковой системы или совокупности взаимодействующих систем. К числу определяющих параметров относятся объем, состав и динамичность информации; число, периодичность и виды запросов; допустимое время ответа на запрос; число абонентов системы и их удаленность от массивов информации. Учитывая перечисленные параметры, выбирают тип ЭВМ и тип запоминающих устройств системы; вид и количество периферийного оборудования; режим работы системы; способы размещения информации в памяти ЭВМ; состав математического (программного) обеспечения системы.

Общими требованиями к техническим средствам со стороны ФИПС являются:

- наличие надежной долговременной памяти, способной длительное время сохранять большие массивы информации;
- наличие достаточно емкого оперативного запоминающего устройства, обеспечивающего выполнение задач обработки и поиска информации; достаточное быстродействие ЭВМ;
- наличие удобных вводных и выводных устройств, в том числе удаленных терминалов, а также устройств, позволяющих выводить информацию в виде таблиц, текста и графиков;
- возможность поиска объектов по любой характеристике (ассоциативный поиск) за сравнительно короткое время;
- возможность работать в режиме многопрограммной пакетной обработки и в режиме диалога.

Существующие в настоящее время ЭВМ и средства их математического обеспечения способны удовлетворить все эти требования. Современный парк ЭВМ составляют машины второго и третьего поколения. Первые из них еще широко эксплуатируются, хотя производство их значительно сократилось; использование машин третьего поколения еще не достигло своего рас-

цвета, но уже появились первые модели машин четвертого поколения. Установлено, что срок службы одного поколения ЭВМ примерно 10 лет, а каждое новое поколение появляется через 6—7 лет [34].

Основные направления прогресса в области электронно-вычислительной техники следующие:

- повышение надежности и уменьшение габаритов ЭВМ;
- увеличение объемов оперативной памяти и внешней памяти прямого доступа;
- увеличение быстродействия;
- возможность работать со словами переменной длины и осуществлять обращение к минимальному объему памяти;
- возможность наращивать центральный процессор и изменять конфигурацию (набор) внешнего оборудования; взаимозаменяемость аппаратуры;
- совершенствование и стандартизация математического обеспечения, совместимость ЭВМ разных моделей на уровне программ;
- возможность работать в мультипрограммном режиме, в системах с разделением времени и диалоговых режимах;
- совершенствование средств ввода, вывода и передачи данных;
- обслуживание абонентов, удаленных от основного процессора;
- создание вычислительных комплексов на базе подключения к универсальной ЭВМ мини-ЭВМ в качестве терминальных устройств.

К универсальным ЭВМ второго поколения относятся БЭСМ-4, БЭСМ-6, М-220, М-222, «Урал-11 и 14», «Минск-22 и 32» и другие машины. Базовой машиной для автоматизированных систем управления в большинстве случаев является «Минск-32», которая по составу математического обеспечения и техническим средствам среди машин второго поколения наиболее приспособлена для информационных задач.

К машинам третьего поколения относятся ЭВМ серии ЕС, разрабатываемые странами—членами СЭВ, а из малых ЭВМ—управляющие машины серии АСВТ (М-5000, М-4030 и др.). Машины этого поколения рассчитаны на многопрограммную и многопроцессорную работу, имеют модульную структуру, развитое математическое обеспечение, переменный набор устройств ввода—вывода, в том числе устройства для телеобработки.

Общая схема многопрограммной ЭВМ указана на рис. 4. Основные параметры универсальных ЭВМ указаны в табл. 8.

Математическое обеспечение ИПС

К математическому обеспечению автоматизированных ИПС относится совокупность программ вместе с их описаниями и

Основные характеристики современных ЭВМ

Виды ЭВМ	Объем ОП (слов)	Объем внешней памяти (слов)		Быстродействие (операция/с)	Режим работы	Наличие системного математического обеспечения	Трансляторы с алгоритмических языков	Минимальная длина слова по обращению (разряды)
		последовательного доступа	прямого доступа					
«Минск-22»	8192	Макс. 1,6 млн.	—	5—6 тыс.	Однопрограммный	Монитор	КОБОЛ, АЛГОЛ-60, ФОРТРАН, АВТОКОД	37
«Минск-32»	16384—65536	Макс. 33 млн.	—	Макс. 65 тыс.	Четырехпрограммный, возможно разделение времени	ОС	КОБОЛ, ФОРТРАН, РПГ, ЯСК	7
БЭСМ-4	8192	4 млн.	МБ 65536	20 тыс.	Однопрограммный	ОС	АЛГОЛ-60, ФОРТРАН, АВТОКОД	45
БЭСМ-6	32768	32 млн.	МБ 362 тыс.	1 млн.	Разделение времени (диалог), пакетная обработка	ОС	АЛГОЛ-60, ФОРТРАН, ЛИСП, АВТОКОД	48
М-220 и М-222	4096—16384	4—16 млн.	МБ 24—576—49152	27 тыс.	Однопрограммный	—	АЛГОЛ-60, ФОРТРАН, АВТОКОД	45
УРАЛ-11 и 14	Макс. 65536	1—48 млн.	МБ 180 тыс.—1 млн. МД макс. 440 тыс.	50 тыс.	Однопрограммный	—	АЛГОЛ-60	46
ЕС-1020	64—256 кб	66 мб	МД 14,5 мб	10—20 тыс.	Пакетная обработка, мультипрограммирование с фиксированным числом задач	ДОС/ЕС	ФОРТРАН (базовый), ПЛ-1, РПГ, КОБОЛ, АССЕМБЛЕР	8
ЕС-1030	128—512 кб	88 мб	МД 21,75 мб	60—100 тыс.	Пакетная обработка, разделение времени, диалоговые режимы, реальный масштаб времени	ДОС/ЕС	То же ПЛ-1, АЛГОЛ, КОБОЛ, ФОРТРАН IV, РПГ, АССЕМБЛЕР	8
ЕС-1040	256—1024 кб	110 мб	МД 36,25 мб	300 тыс.		ОС/ЕС		То же
ЭВМ четвертого поколения (США)	256—1024 кб 512—2048 кб	110 мб 176 мб	МД 36,25 мб МД 43,5 мб	500 тыс. 2 млн. 100 млн.— 500 млн.	Многoprogrammные, многопроцессорные режимы	ОС/ЕС ОС/ЕС	То же >	8 8

Примечание. 1 кб=1024 байт; 1 мб=1 тыс. кб.

инструкциями, обеспечивающих работу ЭВМ и функционирование информационно-поисковой системы.

Математическое обеспечение (МО) принято делить на общее, или машинное, и специальное. Общее МО разрабатывается вместе с определенной моделью ЭВМ и поставляется с ней в одном комплекте. Общее МО позволяет эффективно использовать ЭВМ как универсальную систему обработки информации во всех ее режимах и процедурах. Оно появилось вместе с машинами второго поколения, но в основном получило развитие за последнее время.

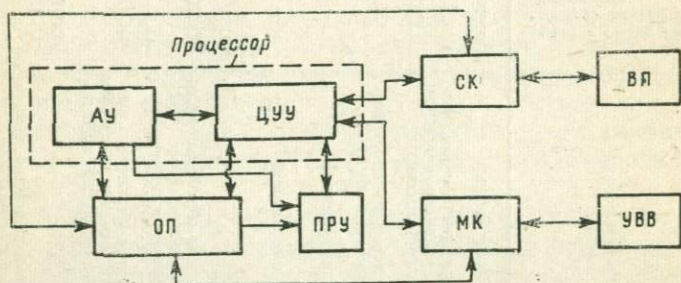


Рис. 4. Основные устройства многопрограммной ЭВМ:

Процессор, включающий АУ — арифметическое устройство и ЦУУ — центральное устройство управления; ОП — оперативная память; ПРУ — пульт ручного управления; СК — селекторные каналы; ВП — внешняя память (на МЛ, МД, МБ); МК — мультиплексный канал; УВВ — устройство ввода-вывода информации (ПК, ПЛ, пишущая машинка, АЦПУ, графические устройства, пульта абонентов и др.)

Специальное МО обеспечивает выполнение функций ЭВМ, для которых предусматривается ее использование (АСУ, ИПС, АСОД и т. д.). Оно разрабатывается самостоятельно. Разработка специального МО представляет собой сложный и трудоемкий этап создания АФИПС. Сложность его зависит от того, насколько развито в ЭВМ системное (общее) МО. В последнее время преобладает тенденция создавать развитое универсальное общесистемное МО. Это намного облегчает труд программистов, так как позволяет генерировать систему в основном за счет программ, имеющихся в комплекте с ЭВМ.

Перечислим основные составляющие элементы МО современных ЭВМ. Подробное их описание приводится в следующих работах [4, 75, 84].

В состав общего математического обеспечения входят:

- операционная система (или набор служебных программ);
- системы программирования;
- программы технического обслуживания и тестовые.

В состав специального МО входят:

- комплект программ по расширению операционной системы и дополнительные средства автоматизации программирования;

— комплект программ, обеспечивающих функционирование ИПС.

Операционная система — это основная часть общего МО ЭВМ, представляющая собой комплекс программных средств, расширяющих аппаратные возможности ЭВМ. Операционная система включает управляющую и обрабатывающие программы. Операционные системы предназначены для работы в разных режимах. Различают однопрограммную работу ЭВМ, при которой обслуживается одна программа (пользователь), и многопрограммную, при которой одновременно обслуживается несколько программ.

Основными режимами работы многопрограммных ЭВМ являются мультипрограммирование — один из видов пакетной обработки задач, характеризующийся одновременным выполнением пакета задач без возможности непосредственного доступа потребителя; параллельная обработка данных, позволяющая осуществлять прямой доступ потребителя к задаче во время ее выполнения; режим разделения времени, имеющий несколько модификаций. Для режима разделения времени характерны мультипрограммирование с непосредственным доступом потребителя и работа в реальном масштабе времени [4].

Операционные системы имеются для машин «Минск-32», БЭСМ-6 и ЕС ЭВМ (см. табл. 8). Для младших моделей ЕС ЭВМ, имеющих малый объем оперативной памяти (64—256 кб) и ограниченный набор внешних устройств, предназначается дисковая операционная система ДОС/ЕС. Для старших моделей ЕС ЭВМ, обладающих объемом памяти свыше 256 кб и полным набором внешних устройств, разрабатывается операционная система ОС/ЕС. По сравнению с ДОС/ЕС ее функции значительно расширены.

Системы программирования входят в состав общего МО. Они состоят из алгоритмических языков и трансляторов к ним, преобразующих программы с языка алгоритмов на язык, понятный машине, т. е. в совокупность машинных команд. Языки программирования будут рассмотрены ниже.

Тестовые программы и программы технического обслуживания используются для отладки программ и обнаружения неисправностей в машине.

Комплекс программ по расширению операционной системы может содержать специальные программы, ориентированные на выполнение специфических задач, например трансляторы со специальных языков, программы работы в определенном режиме, принятом для данной системы, и др.

Комплекс программ по функционированию ИПС должен содержать как программы, связанные с перемещением, поиском и преобразованием данных, так и программы математической обработки информации. Эти программы рабо-

тают под управлением операционной системы. Совокупность программных средств для выполнения определенной функции получила название пакета прикладных программ

В состав специального МО АФИПС входит следующий стандартный набор основных программ:

- ввода и кодирования информации;
- корректировки введенной информации;
- формирования информационных массивов, включая размещение их на машинных носителях и формирование справочных массивов (каталогов);
- сортировки информации для создания и поддержания упорядоченных массивов, а также для выбора информации в соответствии с заданными признаками;
- внесения изменений и дополнений в массивы данных;
- поиска данных в массивах различной структуры;
- сравнения данных с учетом взаимоотношений между единицами информации;
- вывода информации и формирования выходных сообщений;
- обработки запросов;
- решения конкретных задач.

Процедура поиска информации является основной. Она выполняется также в процессе ряда других программ, например в программе внесения изменений, корректировки, сортировки. Поиск может быть двухступенчатый. Сначала находится адрес информации во внешней памяти, а затем поиск и перенос ее в оперативную память для сравнения или преобразования.

Алгоритмические языки

В отличие от языков информационно-поисковых, предназначенных для описания и обозначения семантического содержания информации, языки алгоритмические служат для формального описания процедур, связанных с обработкой этой информации. По алгоритмическим языкам в настоящее время существует обширная литература [например, 87], поэтому укажем только на некоторые их общие особенности. Алгоритмические языки, как и другие языки, имеют алфавит, семантику, задаваемую набором слов, и синтаксис, задаваемый правилами построения выражений. Эти языковые средства позволяют описывать набор данных, процедуры обработки данных и последовательность выполнения процедур.

Описание процедур обработки информации, т. е. собственно программа, может быть записано на разных уровнях детальности. Соответственно алгоритмические языки различают по уровням.

Языком нулевого уровня считается язык кодов машинных команд. Программа обработки информации, составленная на

этом уровне, предписывает выполнение конкретных операций на языке, понятном самой машине. Содержание программы складывается в этом случае из простейших операций с указанием реальных адресов ячеек. Для составления программы на этом уровне нужно учитывать распределение памяти ЭВМ. Процессы составления и отладки такой программы являются трудоемкими.

Во время эксплуатации машин второго поколения получила распространение автоматизация программирования — составление программ на собственно алгоритмических языках с последующей трансляцией их на язык машинных кодов самой машиной. Алгоритмические языки различаются по уровням и по назначению. Языки низшего уровня описывают программу так же подробно, как и языки нулевого уровня, но программа пишется в символических языках с применением мнемонических обозначений операций и констант. Языки этого уровня ориентированы на использование в конкретной модели ЭВМ и поэтому получили название машинно-ориентированных. Во многих машинно-ориентированных языках широко используются также макрокоманды. К языкам этого уровня относятся различные автокоды, язык символического кодирования (ЯСК), ассемблер. Языки низших уровней используются для служебных программ и для составления отдельных подпрограмм. Достоинством их является то, что они позволяют максимально учитывать все особенности машины; программы, составленные на них, отличаются гибкостью.

К языкам высших уровней относят процедурно- и проблемно-ориентированные языки. Программа, составленная на языке высокого уровня, представляет собой описание логической схемы решения задачи — ее алгоритм. Составление такой программы возможно без изучения особенностей конкретной машины и доступно специалистам — постановщикам задач. Трудоемкость составления программы на языке высокого уровня снижается в 5—10 раз по сравнению с составлением программы на языках низших уровней.

В настоящее время языки высокого уровня еще недостаточно стандартизированы и нередко отличаются для различных моделей ЭВМ. Трансляторы с языков различны для каждой модели ЭВМ. Построение транслятора с языка представляет собой достаточно сложную задачу; это ограничивает использование алгоритмических языков. Известно несколько сотен языков программирования, но трансляторы разработаны для меньшей их части. Широкое распространение нашли только некоторые языки.

Наиболее универсальными сейчас признаны языки АЛГОЛ-68 и ПЛ-1, но для первого из них практически еще нет готовых трансляторов.

АЛГОЛ-60 — один из первых универсальных языков про-

граммирования получил широкое развитие у нас в стране и за рубежом для программирования задач вычислительного характера, а также как международный язык публикаций алгоритмов. Входной язык (в отличие от эталонного) различается для разных ЭВМ. Имеется также немало языков программирования, в основе которых используются принципы и правила АЛГОЛ. К ним относятся АЛГЭМ для «Минск-22» и АЛГЭМ для «Минск-32». Данные языки предназначены для обработки информационных массивов, в частности массивов, представленных списковыми структурами. На базе АЛГОЛ-60 построен также язык СИМУЛА-67 и ряд других.

ФОРТРАН (Formula Translation), как можно судить по названию, предназначен для программирования математических задач. Он имеет несколько версий, из которых наибольшее распространение получили ФОРТРАН-IV и его подмножество — базовый ФОРТРАН.

ФОРТРАН — достаточно универсальный язык и может использоваться в различных задачах, но наиболее эффективен он в своей области.

КОБОЛ (Common Business Oriented Language) разрабатывался как язык обработки экономической информации. Основная особенность КОБОЛ — близость его языка к языку естественному, что резко снижает время, нужное для составления и отладки программы. Основным недостатком КОБОЛ — зависимость от типа ЭВМ. В СССР используются преимущественно две версии КОБОЛ — русский вариант для ЭВМ «Минск-32» и КОБОЛ для ЕС ЭВМ, совпадающий с английской версией. КОБОЛ имеет целый ряд преимуществ для обработки информации, обеспечивая удобный способ описания данных с заданием их формата, отделение в программе описания данных от процедур их обработки, возможность подключать подпрограммы, написанные на других языках.

КОБОЛ является наиболее удобным языком для обработки информации, имеющей сложную иерархическую структуру, однако трансляторы с этого языка имеются не для всех ЭВМ (см. табл. 8).

ПЛ-1 разработан фирмой ИБМ. При его создании использованы положительные качества трех языков: АЛГОЛ-60, ФОРТРАН и КОБОЛ. Транслятор с ПЛ-1 входит в состав системного математического обеспечения ЭВМ серии ЕС. Этот язык может использоваться для широкого класса задач и при работе ЭВМ в разных режимах.

РПГ (Report Program Generator) является языком высокого уровня, предназначенным для обработки данных и представления результатов в виде различных отчетов, таблиц, справок. При программировании задач на этом языке описываются структура входных документов, вид процедуры преобразования данных и форма результата. Используется он в основном для

бухгалтерских и плановых задач и может применяться при составлении сводок по фактическим данным.

Для обработки данных, имеющих списковую и нестандартную структуру, кроме рассмотренных универсальных языков, имеющих для этого особые средства, используется целый ряд специальных языков. К ним относятся ЛИСП, КОМИТ, СЛИП и др.

Глава 6

СПОСОБЫ ОРГАНИЗАЦИИ ИНФОРМАЦИИ В ПАМЯТИ ЭВМ

Виды информационных массивов

Информация, накапливаемая в памяти ИПС, фиксируется в запоминающих устройствах ЭВМ, образуя информационные массивы*. Объем информационных массивов намного превышает объем оперативной памяти, поэтому они располагаются на внешних накопителях и по мере необходимости порциями вводятся в оперативную память, где подвергаются различным преобразованиям.

К основным видам преобразований информации относятся структурные преобразования, в том числе выборка, поиск, сортировка, корректировка; математическая обработка — выполнение арифметических операций и решение задач с помощью различных математических методов; перемещение — ввод и вывод информации.

Названные преобразования, составляющие основу функционирования ИПС, во многом зависят от способа организации информации, т. е. от способов размещения ее на различных уровнях памяти и от того, как обеспечивается доступ к этой информации.

Критериями оптимальной организации информации в ИПС обычно считаются такие факторы, как минимальное время поиска, производительность системы (число ответов, которые она может выдать за единицу времени) и экономное расходование быстродействующей памяти.

По функциональному принципу информационные массивы ФИПС можно разделить на:

- главные, содержащие основной комплект фактической информации;
- вспомогательные, или справочные, обеспечивающие функционирование ИПС;
- рабочие массивы.

Рабочие массивы являются временными; они создаются для решения определенных задач и состоят из целенаправленной

* Информационные массивы называют также файлами и наборами данных.

выборки информации из главного массива или новой информации, которая еще не внесена в основной массив. Для главных и справочных массивов АФИПС, как правило, необходимо осуществлять процесс их постоянного ведения, т. е. вносить изменения и корректировать информацию.

Справочные массивы формируются для оптимизации поиска данных в главных массивах, а также для кодирования и декодирования сообщений. В этих массивах обычно содержатся сведения следующего характера: 1) некоторая фактическая информация относительно основных объектов АФИПС, упорядоченная таким образом, что структура записей этой информации не совпадает со структурой записей в главном массиве. Примером могут служить инверсные массивы, в которых информация соотнесена или упорядочена применительно к какому-либо определенному признаку или значению признака; 2) сведения, указывающие на наличие определенной информации в главном массиве. К ним относятся логические шкалы, позиционные матрицы; 3) тезаурус системы — словари, указатели смысловых связей между терминами, правила сравнения и критерии соответствия; 4) информация о размещении фактических сведений на физических носителях памяти (каталоги, указатели).

Всю информацию об объектах в АФИПС с точки зрения поиска можно также подразделять на собственную информацию и ассоциативную, или списковую [40]. Собственная информация находится в главных информационных массивах, она не выносится в качестве поисковой и обнаружить ее можно только прямым считыванием информации об объекте. Ассоциативная информация увязана между собой адресами связи так, что каждый элемент информации содержит сведения о следующем аналогичном элементе информации или об объекте, к которому эта информация относится. Организация информации с использованием адресных ссылок, позволяющая проводить поиск по признакам, получила название методов организации списков. В принципе можно организовать главный информационный массив АФИПС таким образом, чтобы не было различий между списковой и собственной информацией об объекте и вся информация прослеживалась бы цепочками адресов, как и ассоциативная. Однако этот способ требует большого дополнительного расхода памяти и использование его для всей информации может оказаться неэкономичным.

Ассоциативная информация должна храниться в виде поисковых деревьев или ассоциативных узлов [40]. Она может записываться отдельно от собственной информации или совместно.

Состав и структура главных и справочных массивов, а также их размещение на определенных ЗУ определяется в каждом случае самостоятельно в зависимости от ряда факторов.

Структура информационных массивов

Информационный массив представляет собой наиболее крупную логически связанную структурную единицу информации, состоящую из множества логических единиц информации одного вида, называемых записями. Записи, в свою очередь, состоят из наиболее мелких единиц информации, которые можно назвать элементами. На содержательном уровне информационный массив фактографической ИПС — это совокупность сведений о некотором множестве объектов. Запись — это совокупность сведений об одном объекте. Элементом информации является одно свойство или одна из характеристик объекта. Элементы информации могут быть представлены символом, числом или словом в зависимости от способа и алфавита кодирования. Как правило, целесообразно большинство элементов информации записывать в условных числовых кодах, закрепленных за каждым из них: имена собственные представлять в текстовой форме (т. е. в побуквенном коде), а вспомогательные термины и отношения между терминами — в виде символов.

Записи могут быть постоянной, переменной или неопределенной длины в зависимости от состава и числа элементов. Переменная и неопределенная длина записи обусловлена тем, что некоторые характеристики объектов могут иметь по несколько значений и не являться взаимоисключаемыми, а также тем, что ряд сведений может иногда в записи отсутствовать.

Размеры информационных массивов, записей и элементов измеряются единицами физической памяти. К последним относятся ячейки*, части ячеек (биты, семиразрядные символы, восьмиразрядные байты и т. д.), зоны магнитной ленты (МЛ), дорожки магнитных дисков (МД), катушки МЛ, диски и пакеты дисков. Максимальная физическая единица информации называется томом. Она может быть представлена, например, катушкой МЛ, пакетом дисков.

Информационные массивы могут состоять из одинаковых по своей структуре записей, относящихся к объектам одного типа, и неодинаковых, относящихся к разным объектам. В АФИПС целесообразно сортировать записи и включать в один массив только однородные. Отдельные информационные массивы могут увязываться при этом, если необходимо, общими элементами информации.

Типы записей и их расположение в информационном массиве образуют его структуру. По способу расположения записей можно различать четыре основных типа структур информационных массивов 1) последовательные, 2) списковые, 3) древовидные, 4) табличные. Эти типы структур во многом обуслов-

* Ячейки имеют неодинаковую разрядность в разных ЭВМ. В машинах серии ЕС длина ячеек может изменяться: 8, 16, 32, 64 битов.

лены характером самой информации и ее предмашинным представлением. С точки зрения формы информации до ввода ее в ЭВМ может быть запись трех типов: 1) линейная в виде строчек, 2) табличная и 3) иерархическая (древовидная). В памяти ЭВМ вся информация записывается линейно. Основное отличие состоит в том, что логические записи могут либо пространственно контактировать друг с другом, либо быть разобщенными. В связи с этим можно выделять два типа структур — последовательные и списковые, так как табличные и древовидные являются частным случаем тех и других или их комбинацией.

Последовательные информационные массивы

Основной особенностью этих массивов является то, что в них логические последовательности записей размещаются также последовательно и в физическом отношении, т. е. в порядке их внесения в массив. Записи в последовательном массиве могут быть упорядоченными и неупорядоченными. Если массив упорядочен по некоторому признаку, который называют иногда ключом сортировки, то поиск единичных записей по этому ключу в данном массиве осуществляется намного быстрее, чем в неупорядоченном массиве. Поиск в неупорядоченном массиве может проводиться только путем последовательного перебора каждой записи до встречи с запрашиваемой. Если массив большой, то на эту операцию может потребоваться довольно много времени. Поэтому чаще всего предпочитают иметь упорядоченные массивы. Упорядочивание массивов происходит в результате сортировки информации. Существует значительное число различных методов сортировки информации, записанной как в оперативной памяти — внутренняя сортировка, так и во внешней памяти — внешняя сортировка [50, 52, 65].

Поиск информации в упорядоченном массиве может проводиться несколькими методами. Одним из наиболее распространенных является двоичный, или логарифмический, поиск. Любая запись при использовании этого способа может быть найдена за $\log_2 n$ шагов, где n — число записей в массиве [50].

Однако в ряде случаев, когда происходит частое изменение в массиве, требующее вставления или исключения записей, целесообразнее иметь неупорядоченные массивы, так как вставление записи в упорядоченный массив связано с передвижением в среднем половины всех записей. Недостатки упорядоченных массивов обнаруживаются также при многоаспектном поиске, так как обычно массив упорядочивается только по одному признаку.

Основное преимущество последовательных информационных массивов — экономное расходование памяти, а для неупорядоченных массивов — легкость внесения новых записей. При поиске информации по запросу, который требует просмотра

всех или большинства имеющихся в массиве записей, этот способ организации массивов является наиболее удобным.

По содержанию записей существует два способа построения последовательных информационных массивов — прямой и инверсный. При прямом способе записывается сначала номер или наименование объекта, а затем все характеристики этого объекта. При инверсном способе сначала записывается название или значение характеристики, а затем следуют номера объектов, отвечающих этим значениям.

В фактографической системе главный информационный массив удобнее и экономичнее записывать прямым способом, позволяющим извлечь при одном обращении к внешней памяти всю информацию об объекте. Инверсные массивы целесообразно использовать в качестве справочных.

Списковые информационные массивы

К информационным массивам со списковой структурой относятся массивы, в которых размещение записей в физическом отношении не зависит от их логического порядка. Отдельные записи или элементы записей в таких массивах увязаны между собой адресными ссылками, образуя списки и списковые струк-

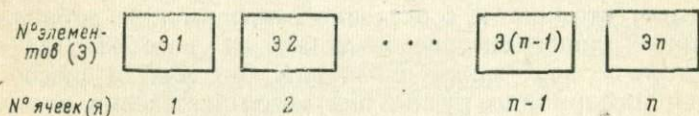


Рис. 5. Последовательный список

туры. Обычно рассматриваются четыре основных способа построения списков: последовательный, цепной, гнездовой, узловый [40]. Первые три способа увязывают между собой элементы (записи) по какому-либо одному признаку или значению признака. Узловой способ позволяет строить сложные списковые структуры, отражающие многоаспектные связи между объектами.

Последовательный список — это частный случай последовательной организации информационных массивов. Так как в этой структуре n -запись следует за $(n-1)$, то нет необходимости проставлять адреса связи (рис. 5).

В цепном списке отдельные элементы (записи) связаны между собой адресами связи, а размещение их на физическом носителе произвольно (рис. 6).

Недостатком этого способа организации массивов является дополнительный расход памяти на проставление адресов связи, зато при доработке и изменении списка корректируются только адреса связи. При организации информации в памяти

ЭВМ цепным способом возможно как совместное, так и раздельное хранение списковой информации и собственной. При раздельном хранении каждый элемент информации из списка сопровождается двумя адресами, из которых один указывает

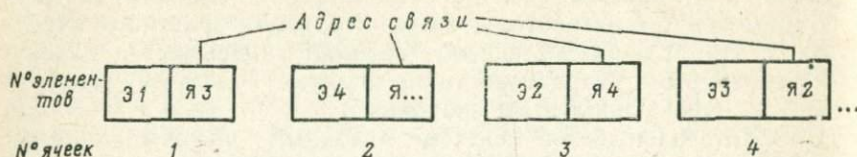


Рис. 6. Цепной список

на адрес следующего элемента списка, а другой — на место хранения собственной информации, к которой имеет отношение данный элемент списка:

Списковое слово	Адрес собственной информации об объекте	Адрес следующего элемента списка
-----------------	---	----------------------------------

Адресная отсылка к собственной информации добавляется также при организации списков любым из способов. Для поиска объектов по нескольким признакам при цепном способе организации информации нужно иметь соответственно несколько цепных списков.

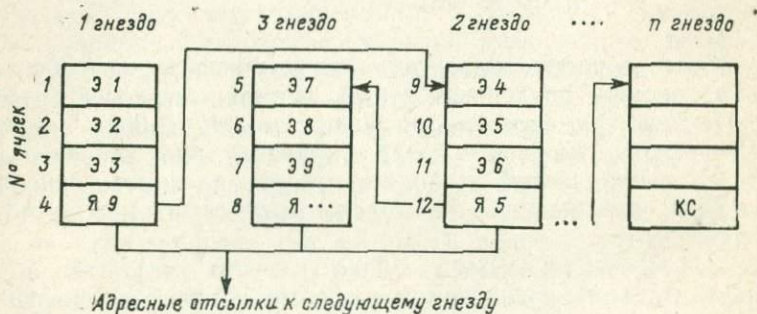


Рис. 7. Гнездовой список:
КС — символ конца списка

Гнездовой список сочетает в себе два предыдущих способа. Часть элементов информации располагается в последовательных ячейках — гнездах, последняя из ячеек гнезда содержит адресную отсылку к следующему гнезду (рис. 7). Гнездовой способ образования списков используется как для

хранения списковой информации, так и собственной информации об объектах. Этим же способом можно вести учет свободных ячеек памяти, что необходимо при организации списковых структур.

Узловой список (мультисписок) позволяет собрать в один узел всю списковую информацию об объекте. В составе узла имеется следующая информация, заголовок узла, наименования и адреса членов списка, дополнительная информация. В качестве заголовка узла может использоваться

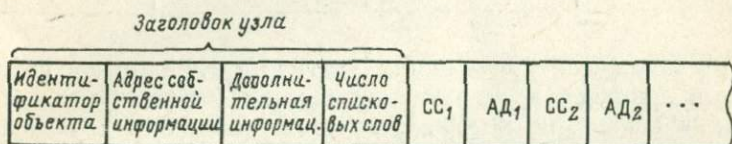


Рис. 8. Строение узла
 СС — списковые слова; АД — адреса следующих элементов списка

машинный идентификатор объекта или его адрес. К дополнительной информации относится адрес собственной информации об объекте (при раздельном хранении), число списковых слов в узле и некоторые другие данные (рис. 8).

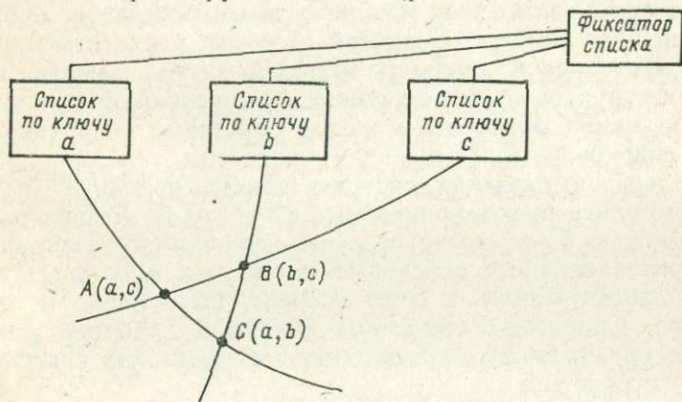


Рис. 9. Узловой список (мультисписок):
 А, В, С — узлы

При узловом способе организации массивов каждый узел, соответствующий определенному объекту, представляет собой как бы место пересечения последовательных или цепных списков с разными ключевыми словами. Схематически это показано на рис. 9. Максимальное число списковых слов для каждого объекта равно числу списков. Для каждого из списков имеется фиксатор или иначе справочник ключей, представляющий собой ячейку или группу ячеек памяти, в которых указаны код ключа

ческого слова, адрес первого элемента списка узла, число элементов в списке и другие данные, связанные со структурой узлов.

Фиксатор списка

Ключевое слово	Адрес первого элемента списка	Число элементов в списке	Дополнительная информация
----------------	-------------------------------	--------------------------	---------------------------

При поиске по конъюнкции ключевых слов по фиксаторам, которые хранятся в оперативной памяти, выбирается наиболее короткий список, по которому ведется поиск. Это делается для уменьшения числа обращений к внешней памяти.

В работе Д. Лефковица [53] приводится сравнение поисковой процедуры по конъюнкции ключей для двух способов организации списков, применяемых для информации одного и того же вида, — узловых списков и последовательных инвентаризованных.

Основным преимуществом мультисписков является программная простота реализации и гибкость обновления; отрицательным фактором является необходимость неоднократного обращения к внешней памяти. С другой стороны, недостатком инвертированных списков является большой расход памяти и программные трудности, возникающие при необходимости упорядочивать список объектов, а также связанные с переменной длиной списков.

С целью оптимизации системы поиска предложены модификации мультисписков, описанные в работе Д. Лефковица [53].

С помощью перечисленных выше способов образования списков строятся сложные списковые структуры, в которых любой элемент списка может, в свою очередь, быть началом любого подсписка. Списковые структуры являются наиболее удобным способом организации данных, имеющих сложную классификационную структуру.

Информационные массивы в виде деревьев

Разветвляющиеся структуры информационных массивов типа деревьев, являющиеся частным случаем списковых структур, имеют широкое применение и значительную перспективу для создания массивов фактографической информации.

Особенностью структуры этого типа является наличие некоторой начальной, не имеющей входа вершины, называемой корнем, и совокупности вершин с одним входом и несколькими выходами.

Общий вид структуры типа «дерево» показан на рис. 10, а, его реализация в виде списковой структуры — на рис. 10, б. С помощью структуры типа «дерево» обычно организованы поисковые массивы и словари. В том случае, когда собственная информация размещается вместе с ассоциативной, информационные массивы в виде дерева являются главными.

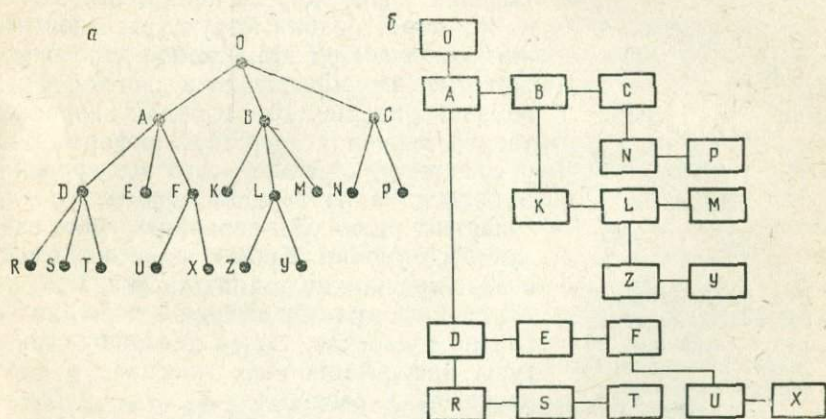


Рис. 10. Дерево (а) и его реализация в виде списковой структуры (б)

Наиболее простым случаем являются двоичные деревья, в которых от каждой вершины отходят две дуги и она содержит информацию о связанных с ней двух вершинах. Обычно правая дуга соединяет с вершиной, имеющей большее значение, чем исходная, а левая дуга — с наименьшим значением вершины. На рис. 11 показан принцип построения двоичного поискового дерева для последовательности значений некоторого признака, имеющей вид

2, 4, 8, 9, 6, 3, 11, 7, 1, 25, 14, 19, 5, 10, 12
 (1) (2) (3) (4) (5) (6) (7) (8) (9) (10) (11) (12) (13) (14) (15)

С помощью такого дерева можно довольно быстро найти любое значение признака, а также достроить дерево любым значением. Алгоритм поиска заключается в последовательном сравнении заданного значения с вершинами двоичного дерева, начиная от корня. Если при сравнении окажется, что n_i заданное $< n_j$ вершины, то дальнейшее сравнение идет по левой дуге, если n_i заданное $> n_j$ вершины, то по правой дуге.

Время поиска по двоичному дереву значительно меньше, чем в неупорядоченном последовательном наборе значений. Минимальное время поиска, равное K (число циклов) $= 1,44 \log_2 N$, где N — число вершин — достигается при сбалансированном де-

реве, у которого путь до каждой концевой вершины отличается не более чем на 1. Однако построение сбалансированных деревьев более сложное, так как требуется первоначальное упорядочение всей последовательности значений.

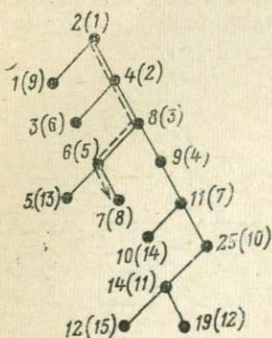


Рис. 11. Двоичное дерево. Путь до находящейся вершины «7» указан пунктиром. Он состоит из четырех отрезков

С. Н. Селеткова и Б. Г. Волкова [8, 57, 64, 72].

Информационные массивы табличной структуры

Информация, представленная в предмашинном виде объектно-характеристическими таблицами, может сохранить эту структуру и в информационных массивах. Это означает, что между позицией матрицы и адресом расположения этой позиции в памяти сохраняется определенная зависимость. Общая структура ОХТ была указана выше. В памяти ЭВМ ОХТ целесообразно хранить в виде трех последовательных списков [72]: 1) наименований объектов, 2) наименований характеристик и 3) строк со значениями характеристик.

Список наименований объектов служит для определения номера строки матрицы любого из объектов и вычисления ее адреса. По списку наименований характеристик определяется адрес значений этой характеристики для каждой строки. Список строк со значениями характеристик является основным массивом данным. В ОХТ с постоянными параметрами, когда число признаков во всех строках одинаково, положение каждой позиции матрицы определено, т. е. известно, к какой характеристике и к какому объекту относится данное значение. Это позволяет проводить поиск в ОХТ также и по значению признака. Однако по сравнению с прямым поиском по наименованию объекта и по наименованию характеристики поиск по значению осуществляется медленнее, так как для этого нужно последовательно просматривать все строки матрицы.

Другим недостатком ОХТ является то, что заранее нужно резервировать часть памяти для размещения строк, что не всегда удобно, особенно, если информация поступает не сразу, а через определенные промежутки времени. ОХТ целесообразно использовать при высоком коэффициенте заполнения.

Известно три способа, позволяющие увеличить коэффициент заполнения: 1) составление логических шкал, 2) деление всей памяти, отведенной под ОХТ, на основную и резервную, 3) деление информации на группы по принципу вероятности присутствия их в описании объекта и раздельное хранение каждой группы [72].

В объектно-характеристических таблицах с переменными параметрами, имеющих разное число характеристик у объектов, соответствие между адресом значения характеристики и номером объекта и признака нарушается. Здесь для реализации поиска требуется также указание адресов хранения для каждого объекта и для каждого значения признака, т. е. в этом случае применяется уже типичная списковая организация информации.

Особенности организации информационных массивов на ЗУ последовательного и прямого доступа

Информационные массивы на магнитных лентах

Магнитные ленты (МЛ) предназначены для длительного хранения информации большого объема. Это наиболее дешевые ЗУ, не считая перфокарт и перфолент, имеющих довольно большое время выборки. Доступ к информации, записанной на магнитной ленте, может быть только последовательным. Для того чтобы осуществить чтение или запись отдельного участка МЛ, нужно прогнать ленту от начала до нужного места. Пересылка информации, записанной на МЛ, в оперативную память осуществляется блоками, или зонами. Записи возможны постоянной, переменной и неопределенной длины. Для записей постоянной и в меньшей мере переменной длины обычно применяется блокирование. Блокирование записей позволяет уменьшить время обращения к МЛ и экономить память за счет промежутков между блоками. Магнитные ленты имеют специальные метки для идентификации ленты, а также служебные знаки, указывающие дату заполнения МЛ, количество записей, срок хранения и другие данные. Метки ставятся в начале и конце магнитной ленты.

Информационные массивы на МЛ записываются последовательно с упорядоченным или неупорядоченным расположением записей. Поиск и выборка информации в неупорядоченных информационных массивах может осуществляться с помощью справочных массивов, устанавливающих соответствие между

определенными признаками и местом хранения описаний объектов с этими признаками на МЛ. При этом указываются зона МЛ и номер записи. Такой справочный массив записывается в начале и (или) конце МЛ либо на устройствах прямого доступа. Он может быть организован любым из описанных выше способов ассоциативного программирования.

В то же время нужно учитывать, что обращение к МЛ по адресу после нахождения этого адреса в справочном массиве не избавляет от прогона МЛ и может оказаться менее эффективным, чем последовательный просмотр записей от начала МЛ. Это связано с тем, что время пустого прогона МЛ через оперативную память ненамного больше времени считывания информации.

Оптимизировать поиск информации, записанной на МЛ, можно путем уменьшения числа обращений к ней. Один из таких способов заключается в подборе наиболее благоприятной стратегии обращений по группе признаков [41]. Получение более быстрого ответа на запрос к информации, записанной на МЛ, достигается также наличием нескольких дублей информационных массивов [65].

Информационные массивы на ЗУ прямого доступа

Особенностью организации данных на ЗУ прямого доступа (ЗУПД) является возможность непосредственного обращения к любому элементу информации, если известен его адрес. Это позволяет эффективно использовать в данном случае ассоциативные методы программирования.

К ЗУПД относятся магнитные диски, магнитные барабаны и магнокорды. В ЕС ЭВМ в качестве ЗУПД используются пакеты сменных магнитных дисков по шесть штук в каждом пакете, с десятью рабочими поверхностями. На каждой поверхности имеется 203 концентрические дорожки (0—199 — рабочие, 200—202 — запасные). Совокупность дорожек, лежащих одна под другой на различных поверхностях дисков, образует цилиндр. Соответственно имеется 203 цилиндра по 10 дорожек в каждом. На дорожках записываются фактические данные и служебная информация, позволяющая фиксировать место каждой записи.

В составе стандартного математического обеспечения ЕС ЭВМ предусматривается четыре возможных способа организации информации: последовательный, библиотечный, индексно-последовательный, прямой. Для каждого из них выполняются либо последовательная выборка, либо прямая — по адресу или ключу. Эти способы описаны К. Джермейном [25], а также в работе «Система ИБМ/360» [74].

Пример организации информации для некоторых геологических объектов

Рассмотрим примерный состав и структуру информационных массивов для хранения данных по некоторым геологическим объектам. Для примера возьмем сложный объект типа месторождений полезных ископаемых. Описание таких объектов в предмашинном формате содержит наименования объектов и набор различных характеристик, причем последние могут быть представлены совокупностью терминов. Общая длина записи сведений об одном объекте точно не определена.

Для подобных геологических объектов целесообразно иметь следующий набор массивов:

- 1) справочный массив наименований объектов;
- 2) массивы ассоциативной информации об объекте;
- 3) главный массив, разделенный на подмассивы с постоянной и переменной длиной записей;
- 4) логические шкалы для указания наличия или отсутствия ряда сведений;
- 5) словарь наименований характеристик;
- 6) словарь дескрипторов и их парадигматических связей;
- 7) словарь понятий, представленных семантическим кодом;
- 8) словарь словоформ (или основ слов и окончаний) для автоматического кодирования.

Функции перечисленных массивов следующие:

1. Справочный массив наименований объектов в общем случае содержит словарные наименования объектов, машинные идентификаторы объектов, начальный адрес описаний объектов в главном массиве и длину каждой записи

Наименование объекта	Идентификатор объекта	Адрес начала записи	Длина записи (в ячейках)
----------------------	-----------------------	---------------------	--------------------------

Этот массив используется для: 1) установления присутствия определенного объекта в массиве (по названию); 2) нахождения его адреса; 3) замены имен объектов на числовой или буквенный идентификатор, который имеет постоянную длину для всех объектов и используется взамен наименований в главном и ассоциативных массивах с целью экономии места и удобства обработки информации. В качестве буквенного идентификатора может быть свертка наименования, а в массивах с постоянной информацией в качестве идентификаторов можно использовать непосредственно адрес объекта в памяти; 4) для декодирования при выдаче сообщений.

Массив наименований можно организовать в виде лексикографически упорядоченного списка (если новые объекты

появляются сравнительно редко); поискового дерева (рис. 12, а, б), если число наименований порядка десятков тысяч и появление новых объектов происходит довольно часто; неупорядоченного перечня, если число наименований объектов сравни-

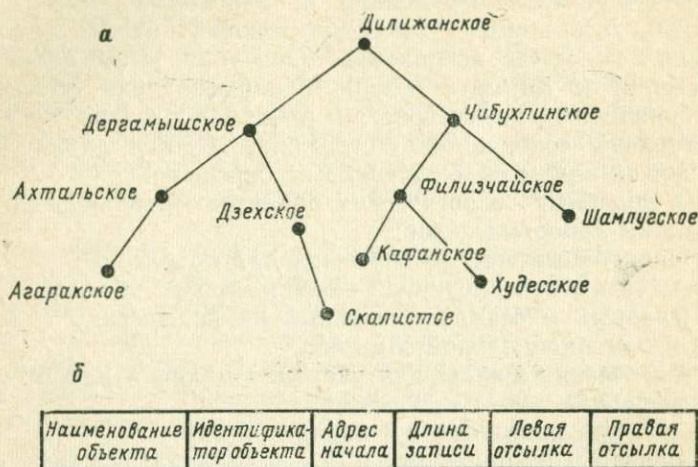


Рис. 12. Двоичное поисковое дерево по наименованиям месторождений (а); строение вершины этого дерева (б)

тельно невелико, а смена наименований происходит довольно часто.

2. Массивы ассоциативной информации для поиска по признакам строятся либо по принципу инверсных массивов, либо способом построения цепных и узловых списков в зависимости от общего числа признаков, которые выносятся в качестве ассоциативной информации, и от вида самих признаков. Например, ассоциативный массив для поиска по территориальному признаку или

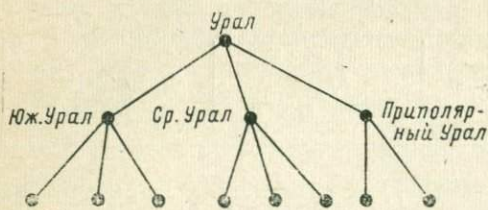


Рис. 13. Разветвляющееся дерево для поиска данных иерархической структуры

другой информации с иерархической структурой целесообразно организовать по принципу разветвляющегося дерева (рис. 13). В каждой вершине содержатся машинный идентификатор объекта и отсылки к другим объектам по принципу, указанному на рис. 10, б.

3. Главный массив размещается в ЗУ таким образом, чтобы за минимальное число обращений к внешней памяти можно было получить всю информацию об одном объекте. Массив мож-

но разделить на два-три подмассива, где будут храниться записи, имеющие примерно одинаковую длину. В одном из них могут быть записаны характеристики, имеющие постоянную длину записей. Подмассивы можно организовать по принципу частоты обращений к отдельным характеристикам объектов. Целесообразно также объединить в одном подмассиве признаки вещественного состава, учитывая как неопределенную длину записи этих данных, так и удобство их сравнения.

Описания всех данных по объекту, размещенных в разных подмассивах, имеют общий идентификатор объекта и увязаны между собой адресными ссылками. Организация записей внутри подмассива последовательная.

4. Логические шкалы можно использовать для указания о наличии или отсутствии данных по объектам и определения порядка расположения характеристик в подмассивах основной информации в качестве грамматических средств.

5. Словарь наименований характеристик используется как справочный для указаний всех видов данных, которыми описываются объекты. Кроме того, по нему можно определить последовательность размещения характеристик в описании объекта и их адрес в ЗУ. Он имеет следующую структуру записи:

Наименование характеристики	Идентификатор характеристики	Место в логической записи	Номер подмассива и его адрес
-----------------------------	------------------------------	---------------------------	------------------------------

Идентификатор наименования характеристики используется в тех случаях, когда свойства объекта записываются в произвольной последовательности, например, при формировании массива изменений, в запросах, при подготовке выходных сообщений.

6. Словарь дескрипторов (терминов) — значений характеристик используется для их кодирования, т. е. замены словарных наименований условными цифровыми кодами. В словаре указываются также парадигматические связи между терминами.

7. Записи в словаре семантических кодов имеют следующий вид:

Наименование понятия	Смысловой код	Условный код
----------------------	---------------	--------------

Словари семантических кодов, являющихся «стандартами» содержания понятий, используются для автоматического смыслового кодирования наименований и декодирования семантических кодов. С их помощью можно присвоить наименование не-

которому понятию по его семантической записи по принципу максимального сходства с некоторым «стандартным». С целью экономии места в записи, а также во всех случаях, когда смысловые коды не нужны, наименования или смысловые коды заменяются условными кодами.

8. Словари словоформ (или словарь основ) и наборы грамматической информации служат для автоматического кодирования. Построение таких словарей специально рассматривается в работе Г. Г. Белоногова и В. И. Богатырева [7].

СПИСОК ЛИТЕРАТУРЫ

1. Абрамович И. И., Груза В. В. Фациально-формационный анализ магматических комплексов. Л., «Недра», 1972, 236 с.
2. Автоматизированные системы информационного комплекса в геологии. М., «Недра», 1973, 199 с.
3. Автоматизированная система поиска и обработки фактографической информации по геологии нефти и газа. — «Тр. ВНИГРИ», 1972, вып. 312, 148 с.
4. Алферова З. В., Лихачева Г. Н., Шураков В. В. Математическое обеспечение ЭВМ. М., «Статистика», 1974, 383 с.
5. Аронов В. И., Невельская Э. Я. Реализация ИПС в геологии на ЭЦВМ. — В кн.: Математические методы и ЭЦВМ в геологии. 1971, 224 с. (Тр. ВНИГРИ, вып. 103).
6. Бейли Н. Математика в биологии и медицине. М., «Мир», 1970, 326 с.
7. Белоногов Г. Г., Богатырева В. И. Автоматизированные информационные системы. М., «Советское радио», 1973, 328 с.
8. Берзтисс А. Т. Структуры данных. М., «Статистика», 1974, 408 с.
9. Бриллюэн Л. Научная неопределенность и информация. М., «Мир», 1966, 271 с.
10. Бугаец А. Н., Мацак А. П. Математические методы при поисках месторождений полезных ископаемых и крупномасштабных металлогенических исследованиях. — В кн.: Применение современных мат. методов и ЭВМ в области геол. в странах — членах СЭВ. Алма-Ата, 1973, с. 61—78.
11. Вакорина Л. И., Рудник В. А. О методике сбора и систематизации первичной петрохимической информации. — В кн.: Вопросы петрохимии (материалы совещания). Л., 1969, с. 57—58 (ВСЕГЕИ).
12. Вакорина Л. И., Лови Б. И., Рудник В. А. Методические основы создания автоматизированной системы сбора, хранения, поиска и обработки петрохимической информации. — «Советская геология», 1973, № 10, с. 151—153.
13. Вернадский В. И. Избранные сочинения. Т. 5, Биосфера. М., Изд-во АН СССР, 1960, с. 17—20.
14. Воронин Ю. А. Введение мер сходства и связи для решения геолого-геофизических задач. — «Докл. АН СССР», 1971, т. 199, № 5, с. 1011—1014.
15. Воронин Ю. А. и др. Геология и математика. Новосибирск, «Наука», 1967, 253 с.
16. Воронин Ю. А., Еганов Э. А. Универсальная схема аналитического описания сложных геологических тел. — «Тр. СНИИГГИМС», 1968, вып. 79, с. 21—35.
17. Воронин Ю. А., Еганов Э. А. О процедурах сопоставления сложных геологических тел на основе их аналитического описания. — «Тр. СНИИГГИМС», 1968, вып. 79, с. 36—48.
18. Геологический словарь. Т. I, 402 с.; Т. II, 455 с. М., Госгеолтехиздат, 1965.
19. Геологический словарь. Т. I, 486 с.; Т. II, 456 с. М., «Недра», 1973.

20. Глушков В. М. Основные принципы построения автоматизированных систем управления. Тезисы докл. Киев, 1969, 34 с.
21. Глушков В. М. Введение в АСУ. — «Техника», 1972, 310 с.
22. Гольдина Н. А., Новикова Р. Г., Суровцева М. Ю. Опыт создания и эксплуатации фактографических ИПС по геологии нефти и газа Западно-Сибирской провинции. — В кн.: Информационные системы. Вып. 1. Киев, 1973, с. 53—54.
23. Гольдшмидт В. И. Классификация геологических объектов по комплексу признаков (комплекс). — В кн.: Алгоритмы и программы для обработки геол.-геофиз. данных на ЭВМ «Минск-2 (22)». Алма-Ата, 1973, с. 6—25.
24. Гусельников И. И. Универсальный определитель минералов на перфокартах краевой перфорации. М., ВЗПИ, 1966, 7 с.
25. Джермейн К. Программирование на IBM/360. М., «Мир», 1973, 870 с.
26. Драгунов В. И. Учение о формациях. — «Тр. ВСЕГЕИ, нов. серия», 1971, т. 177, с. 153—162.
27. Дуденко Л. Н. Теория и практика количественной оценки рудоности территорий. — В кн.: Применение современных мат. методов и ЭВМ в обл. геол. в странах — членах СЭВ. Алма-Ата, 1973, с. 79—83.
28. Дуринов М. А. Информационно-логическая система для решения геологических задач. — В кн.: Применение мат. методов и ЭВМ в геол. Тезисы семинара. Алма-Ата, 1974, с. 227—230.
29. Жуков Р. А., Ткачев Ю. В. Отраслевая автоматизированная система регионального минерагенического прогнозирования (АСУ-Прогноз) и ее роль в АСУ-Геология. — В кн.: Мат. методы в геол., 1973, с. 10—12 (Вестн. Львовск. ун-та).
30. Захаров Е. Е. О некоторых вопросах классификации рудных месторождений. — «Советская геология», 1965, № 9, с. 15—32.
31. Захаров Е. Е., Смирнова А. С. Информационно-поисковый язык фактографический ИПС по рудным месторождениям. — «Научно-техническая информация», 1970, сер. 2, № 12, с. 16—20.
32. Захаров Е. Е., Кристальный Б. В., Смирнова А. С. Об информации, вводимой в автоматизированную фактографическую информационно-поисковую систему (АФИПС) по рудным месторождениям. — «Изв. АН СССР. Сер. геол.», 1970, № 3, с. 82—90.
33. Захаров Е. Е., Смирнова А. С. Схема описания рудного месторождения для фактографической информационно-поисковой системы. — В кн.: Принципы и методы сбора информ. по рудным м-ниям. Тезисы докладов. Л., 1970, с. 36—37 (ВСЕГЕИ).
34. Зейденберг В. К., Матвеевко Н. А., Тароватова Е. В. Обзор зарубежной вычислительной техники по состоянию на 1973 г. М., 1973, 324 с.
35. Иванов С. Н. О характерных особенностях рудных месторождений колчеданного типа. — «Тр. Горно-геол. ин-та Уральск. филиала АН СССР», 1955, вып. 26, с. 86—91.
36. Иванова А. А., Михайлова Ю. И. Принципы сбора и методы регистрации информации по рудным месторождениям при металлогенических исследованиях (на примере мезозойских эндогенных месторождений Вост. Забайкалья). — В кн.: Принципы и методы сбора информ. по рудным м-ниям. Тезисы докладов. Л., 1970, с. 38—39.
37. Игревский В. И. Основные принципы построения автоматизированной системы управления АСУ-Геология. — «Советская геология», 1973, № 1, с. 3—14.
38. Информационные системы. Вып. I, II. Киев, 1973, 72 с., 66 с.
39. Информационно-поисковые системы, реализованные на перфокартах ручного обращения. М., ВИЭМС, 1973, 122 с.
40. Китов А. И. Программирование информационно-логических задач. М., «Советское радио», 1967, 327 с.
41. Кожурин Ф. Л., Ярмош И. А. Структурная обработка больших информационных массивов. Минск, «Наука и техника», 1973, 200 с.

42. Кондаков Н. И. Логика. М., Изд-во АН СССР 1954, 512 с.
43. Константинов Р. М. Основы формационного анализа гидротермальных рудных месторождений. М., «Наука», 1973, 215 с.
44. Константинов Р. М., Дмитриев А. Н. Методы обработки качественной геологической информации для определения формационного типа месторождений. — «Изв. АН СССР. Сер. геол.», 1971, № 8, с. 3—15.
45. К разработке автоматизированных фактографических информационных систем в геологии. — В кн.: Информационные системы. Вып. 1, Киев, 1973, с. 29—33 (Авт.: Г. А. Булкин, К. Б. Гаталин, Т. В. Никифорова, А. Е. Тюленев).
46. Кренделев Ф. П., Дмитриев А. Н., Журавлев Ю. И. Сравнение геологического строения зарубежных месторождений докембрийских конгломератов с помощью дискретной математики. — «Докл. АН СССР», 1967, т. 173, № 5, с. 1149—1152.
47. Крос Р. К., Гардэн Ж. К., Леви Ф. Синтол. Универсальная модель системы информационного поиска. Сб. переводов ВИНТИ, № 10, М., 1968, 178 с.
48. Красный Л. И. Проблемы тектонической систематики. М., «Недра», 1972, 151 с.
49. Круть И. В. Исследование оснований теоретической геологии. М., «Наука», 1973, 203 с.
50. Курбаков К. И. Кодирование и поиск информации в автоматическом словаре. М., «Советское радио», 1968, 247 с.
51. Кухаренко Н. А. Объекты информационных систем по металлогенезу. — «Советская геология», 1974, № 7, с. 119—123.
52. Лавров С. С., Гончарова Л. И. Автоматическая обработка данных, хранение информации в памяти ЭВМ. М., «Наука», 1971, 160 с.
53. Лефкович Д. Структуры информационных массивов оперативных систем. М., «Энергия», 1973, 208 с.
54. Лялько В. И., Шнейдерман Г. А. Методологические принципы создания АСУ-Вода в гидрогеологии. — В кн.: Информационные системы. Вып. 1. Киев, 1973, с. 42—44.
55. Ляцкий В. Б., Гаталин К. Б., Бурдэ А. И. Фактографические информационно-поисковые системы по месторождениям полезных ископаемых как основа для изучения перспективных районов с помощью ЭВМ. — «Научно-техническая информация», 1967, сер. 2, № 12, с. 26—28.
56. Мерзон Л. С. В. И. Ленин о сущности и роли фактов науки. — В кн.: В. И. Ленин и вопросы философской науки. Л., 1970, с. 113—129.
57. Мидоу Ч. Анализ информационно-поисковых систем (введение для программистов). Пер. с англ. М., «Мир», 1970, 368 с.
58. Миронюк Е. П. Фактодокументографическая информационно-поисковая система по картам геологического содержания. — В кн.: Автоматизированные системы информ. комплекса в геол. М., 1973, с. 164—172.
59. Михайлов А. И., Черный А. И., Гиляревский Р. С. Основы информатики. М., «Наука», 1968, 756 с.
60. Михайлова Ю. И., Иванова А. А. О создании комплекса информационно-поисковых систем (РИПКС) для региональных металлогенетических исследований и прогнозирования. — В кн.: Информационные системы. Вып. 1. Киев, 1973, с. 48—50.
61. Мошкин В. Н. Основные научно-технические и организационные предпосылки к созданию автоматизированных информационных систем по геологии. — В кн.: Автоматизированные системы информ. комплекса в геол. М., 1973, с. 5—12.
62. Наливкин В. Д. Границы геологических объектов. — В кн.: Современные проблемы геологии. Т. LXVII, вып. 2, Л., 1974, с. 51—54 (Зап. ЛГИ им. Г. В. Плеханова).
63. Некоторые вопросы комплексирования информационных систем для обеспечения прогнозно-металлогенетических исследований. — В кн.: Информационные системы. Вып. 1. Киев, 1973, с. 50—53 (Авт.: Н. А. Кухаренко, Б. И. Лови, С. А. Топорец, М. В. Унксова).

64. Нечас И. Организация памяти вычислительных машин. М., «Энергия», 1974, 168 с.
65. Обработка информационных массивов в автоматизированных системах управления. Киев, «Наукова думка», 1970, 182 с. (Авт.: В. М. Глушков, В. П. Гладун, Л. С. Лозинский, С. Б. Погребинский).
66. Опыт разработки банков данных для автоматизированной системы прогнозирования. — В кн.: Применение мат. методов и ЭВМ в геол. Тезисы семинара. Алма-Ата, 1974, с. 248—249. (Авт.: Б. А. Чумаченко, Е. П. Власов, Э. А. Немировский и др.).
67. Поваренных А. С. Кристаллохимическая классификация минеральных видов. Киев, «Наукова думка», 1966, 547 с.
68. Применение математических методов и ЭВМ в геологии. Тезисы семинара. Алма-Ата, 1974, 289 с.
69. Применение математических методов и ЭВМ для решения прогнозных задач нефтяной геологии. Кратк. тезисы докл. Новосибирск, 1973, 84 с.
70. Принципы и методы сбора информации по рудным месторождениям. Тезисы докладов. Л., ВСЕГЕИ, 1970, 70 с.
71. Родионов Д. А. Статистические методы разграничения геологических объектов по комплексу признаков. М., «Недра», 1968, 158 с.
72. Селетков С. Н., Волков Б. Г. Хранение и поиск данных в ЭВМ. М., «Советское радио», 1971, 224 с.
73. Сидоров В. М., Воронин Ю. А. Методологические вопросы создания АСУ-Наука в геологии. — В кн.: Информ. системы Вып. 1. Киев, 1973, с. 3—9.
74. Система ИБМ/360. Введение в запоминающие устройства прямого доступа. М., «Статистика», 1974, 125 с.
75. Система математического обеспечения ЕС ЭВМ. М., «Статистика», 1974, 216 с.
76. Скороходько Э. Ф. и др. Информационно-поисковая система БИТ. Киев, «Наукова думка», 1968, 219 с.
77. Смирнов В. А. Уровни знания и этапы процесса познания. — В кн.: Проблемы логики научного познания. М., «Наука», 1964, с. 23—52.
78. Смирнова А. С. О смысловом кодировании некоторых понятий в геологии. — «Научно-техническая информация», 1967, сер. 2, № 10, с. 15—18.
79. Смирнова А. С. О принципах сравнения информации в автоматизированной фактографической ИПС. — «Научно-техническая информация», 1970, сер. 2, № 7, с. 35—37.
80. Смирнова А. С. Примеры некоторых типов задач для фактографической информационно-поисковой системы по рудным месторождениям. Экспресс-информация ВИЭМС. Сер. НТИ в геологии. 1970, 11 с.
81. Смирнова А. С. Разработка автоматизированной фактографической информационно-поисковой системы (на примере ФИПС по рудным месторождениям). М., Науч. собр. ИМГРЭ, 1971, вып. 6, с. 68—78.
82. Тектоническая номенклатура и классификация основных структурных элементов земной коры материков. — «Геотектоника», 1975, № 5, с. 3—21 (Авт.: А. А. Богданов, Л. П. Зоненщайн, М. В. Муратов и др.).
83. Точилин М. С. О диалектическом обобщении в геологии. — В кн.: Философия и естествознание. Вып. 1. Воронеж, 1965, с. 120—150.
84. Трахтенгерц Э. А. Программное обеспечение автоматизированных систем управления. М., «Статистика», 1974, 288 с.
85. Универсальная десятичная классификация. Естественные и технические науки. ВИНТИ, 1965, 138 с.
86. Черный А. И. Интегральные информационные системы (принципы построения, роль и перспективы развития). М., ВИНТИ, 1974, 60 с.
87. Хигман Б. Сравнительное изучение языков программирования. М., «Мир», 1974, 204 с.
88. Черный А. И. Общая методика построения тезаурусов. — В кн.: Науч.-техн. информ. Сер. 2, № 5, 1968, с. 9—32.

89. Четвериков Л. И. Теоретические основы моделирования тел твердых полезных ископаемых. Изд. Воронеж. ун-та, 1968, 152 с.
90. Шатский Н. С. Избранные труды. Т. III. М., «Наука», 1965, 348 с.
91. Швырев В. С. Некоторые вопросы логико-методологического анализа отношений теоретического и эмпирического уровней познания. — В кн.: Проблемы логики научного познания. М., 1964, с. 53—86.
92. Шрейдер Ю. А. Информация и метаинформация. — «Научно-техническая информация», 1974, сер. 2, № 4, с. 3—10.
93. Шрейдер Ю. А. Машинный перевод на основе смыслового кодирования текстов. — «Научно-техническая информация», 1963, № 1, с. 34—38.
94. Шрейдер Ю. А. О количественных характеристиках семантической информации. — «Научно-техническая информация», 1963, № 10, с. 33—38.
95. Anderton M. R. A drillhole data bank. «Appl. Comput. Meth. Miner. Ind. Proc. 10-th Int. Symp., Johannesburg, 1972», Johannesburg, 1973, p. 79—82.
96. Application of data processing in geological mapping in the Grenville project of the Quebec Department of Natural Resources (Abstracts). Bull. Canadian Inst. Min. Metallurgy, 1971, v. 64, № 707, p. 67 (Aut: A. F. Laurin, K. Sharma, A. Franconi and oth.)
97. Blanhet P. H. and Codwin C. I. «Geolog system» for computer and manual analysis of geologic data from porphyry and other deposits. Economic geology, 1972, v. 67, N 6, p. 796—813.
98. Burk C. F. Jr. Development of a national computer-based network of basic information on Canadian mineral deposits. Canadian Min. Jour., 1972, v. 93, N 4, p. 34—38.
99. Burk C. F. Jr. Computer based storage and retrieval of geoscience information: Bibliography 1970—1972; Canadian centre for geoscience Data. Geol. Survey Can. 1973, Paper 73—14, 38 p.
100. Collyer P. L., Merriam D. F. An application of cluster analysis in mineral exploration «J. Int. Assoc. Math. Geol.», 1973, v. 5, N 3, p. 213—223.
101. Cornelius F., Burk C. F. Jr. Toward a national system for geoscience data. Oilweek, 1969, v. 19, № 50, p. 38—40, 48.
102. Cornelius F., Burk C. F. Jr. A petrological-mineralogical code for computer use (Review); Am. Mineralogist, 1971, v. 56, № 3—4, p. 648—49.
103. Derec G., Sarcia I. A. et Troly G. Considerations theoriques sur le traitement de L'information en metallogénie. Chronique des mines. 1964, № 334, p. 251—272.
104. Dimitriu A. L., Dimitriu Cr. Geologic data processing in Romania. «J. Int. Assoc. Math. Geol.», 1973, v. 5, № 3, p. 313—318.
105. Dixon C. J. Semantic symbols. «J. Int. Assoc. Math. Geol.», 1970, v. 2 N 1, p. 81—87.
106. Fabbri A. G. A geological data bank for statistical analysis. «Pap. geol. Surv. Can.», 1973, N 1, part B, p. 29—37.
107. Hawkings Douglas M., Kew Desmond S. Use of multivariate statistical methods in data-base maintenance. «J. Int. Assoc. Math. Geol.», 1974, v. 6, N 1, p. 47—57.
108. Hruška J. and Burk C. F. Jr. Computer based storage and retrieval of geoscience information: Bibliography 1946—1969, Canadian Centre for Geoscience Data. Geol. Survey Can., 1971, paper 71—40, 52 p.
109. Hruška J. Towards computer-processible economic geology data files and their informational aspects. Mineralia slovaca, 1973, v. 5, N 4, p. 453—462.
110. Hubax A. Sheme for a quick description of rocks. «J. Int. Assoc. Math. Geol.», 1971, v. 3, N 3, p. 317—322.
111. Hubax A. A new geological tool—the data Earth. Science reviews, 1973, v. 9, № 2, p. 159—196.
112. Jelinková Jaroslava. Vytvoreni a využití ložiskové Databanky v Severočeskem hnědoudelném reviru. «Symp. pracov. bānsk. pram. Pribram, 1972, Sek. M», S 1, p. 161—178, 403—404.
113. Kelly A. M. Recommended standards for recording the location of

mineral deposits: Canadian Centre for Geoscience Data, Geol. Survey Canad., 1972, paper 72-9, 8 p.

114. Laffitte P. Limites actuelles de l'informatique géologique. «Bull. Bur. rech. géol. et minières». 1968, sec. 4, N 3, 1-9.

115. Laffitte P. La codification sémantique géologique. Ann. mines, 1969, dec., p. 75-83.

116. De Geoffroy J., Wignall T. K. Design of statistical data-processing system to assist regional exploration planning. Part I, «Can. Mining J.», 1973, v. 94, N 11, p. 30-35.

117. Levien R. E., Maron M. E. A computer system for inference execution and data retrieval. Commons. ACM, 1967, v. 10, N 11, p. 715-721.

118. Pamentier C. B. What is the SAFRAS system? Oilweek, 1971, v. 21, N 50, p. 27-28.

119. Perry I. W., Kent A. Tools for machine literature searching. Semantic code dictionary. Equipment, Procedures. N-J. Interscience Publishers, Inc. 1958, 972 p.

120. Ranganathan S. R. Colon Classification. Bombay, 1963.

121. Robinson S. C. Interim report on possible application of data processing techniques to storage and retrieval of geological data in Canada. Canad. Geol. Surv., Jan. 1965, 79 p.

122. Smith James G., Berg Henri C. Use of machine-processable field notes in a wilderness mapping project (Granite Fiords area) southeastern Alaska. J. Res. U. S. Geol. Surv., 1973, v. 1, № 5, p. 579-585.

123. Wynne-Edwards H. R. A geological field data system for the Grenville project in Quebec (Abstr.): Geol. Soc. America. Abstr. with Programs, 1970, v. 2, № 7, p. 729-730.

ОГЛАВЛЕНИЕ

	Стр.
Предисловие	3
Часть I. Некоторые общие вопросы создания фактографических информационно-поисковых систем в геологии	5
<i>Глава 1. Фактографические ИПС и их функции</i>	<i>5</i>
Типы информационно-поисковых систем	5
Предпосылки к созданию автоматизированных ФИПС в геологии	7
Некоторые трудности на пути создания АФИПС	9
Значение АФИПС для создания АСУ и решения геологических задач	10
Составные части АФИПС	11
Состояние разработок АФИПС и примеры некоторых зарубежных систем	13
Часть II. Семантические аспекты построения АФИПС в геологии	17
<i>Глава 2. Методы анализа и способы классификации геологической информации</i>	<i>17</i>
Виды геологической информации и ее источники	17
Способы классификации геологической информации	21
Требования к выбору информации для АФИПС	23
Этапы отбора информации для АФИПС	26
Обмен информацией между АФИПС	27
<i>Глава 3. Информационно-поисковый язык фактографических систем</i>	<i>29</i>
Общие сведения о ИПЯ и формах представления фактографической информации	29
Роль ИПЯ для поиска информации. Модель понимания	32
Составные части ИПЯ	34
Способы формализации геологических понятий. Парадигматика ИПЯ	36
Синтагматика ИПЯ, способы построения высказываний	49
Пример ИПЯ с синтагматическими отношениями типа «стандартных фраз»	53
<i>Глава 4. Методы сравнения информации в фактографической ИПС</i>	<i>55</i>
Общие замечания	55
Установление степени сходства между двумя элементами информации в АФИПС	57
Сравнение объектов информации по комплексу свойств	61
Часть III. Технические средства и математическое обеспечение АФИПС	65
<i>Глава 5. Общие сведения о ЭВМ и математическом обеспечении систем</i>	<i>65</i>
Основные требования к техническим средствам со стороны АФИПС и виды ЭВМ, используемые в настоящее время	65
Математическое обеспечение ИПС	66
Алгоритмические языки	70
	95

Глава 6. Способы организации информации в памяти ЭВМ	73
Виды информационных массивов	73
Структура информационных массивов	75
Последовательные информационные массивы	76
Списковые информационные массивы	77
Информационные массивы в виде деревьев	80
Информационные массивы табличной структуры	82
Особенности организации информационных массивов на ЗУ последовательного и прямого доступа	83
Пример организации информации для некоторых геологических объектов	84
Список литературы	89

Ая Степановна Смирнова

**ПОСТРОЕНИЕ
АВТОМАТИЗИРОВАННЫХ
ФАКТОГРАФИЧЕСКИХ
ИНФОРМАЦИОННО-
ПОИСКОВЫХ СИСТЕМ
В ГЕОЛОГИИ**

Редактор издательства *А. М. Поспелова*

Художественный редактор *В. В. Шутько*

Обложка художника *Г. А. Петрова*

Технические редакторы *В. В. Кириллова, В. В. Соколова*

Корректор *М. П. Курылева*

Сдано в набор 24/III 1976 г. Подписано в печать 24/IX 1976 г. Т-18505. Формат 60×90¹/₁₆. Бумага № 2. Печ. л. 6,0. Уч.-изд. л. 5,75. Тираж 1800 экз. Заказ № 2182/5596-14. Цена 58 коп.

Издательство «Недра», 103633, Москва, К-12, Третьяковский проезд, 1/19.
Московская типография № 6 Союзполиграфпрома
при Государственном комитете Совета Министров СССР
по делам издательств, полиграфии и книжной торговли.
109088, Москва, Ж-88, Южнопортовая ул., 24.

58 коп.

2024

НЕДРА