Записки Ленинградского ордена Ленина, ордена Октябрьской Революции
и ордена Трудового Красного Знамени горного института им. Г.В.Плеханова,
т. 102, 1984

УДК 681.3:553.04

ИНФОРМАЦИОННАЯ СИСТЕМА УНИВЕРСАЛЬНОГО НАЗНАЧЕНИЯ С ЕСТЕСТВЕННЫМ ЯЗЫКОМ ДЛЯ РЕШЕНИЯ ЗАДАЧ УПРАВЛЕНИЯ МИНЕРАЛЬНО-СЫРЬЕВЫМИ РЕСУРСАМИ

А.Н.ЛИВШИЦ, М.А.БОРИСОВА,
В.Н.СМИРНОВ, А.Е.ФАЛЬК

Создание системы государственных кадастров месторождений полезных ископаемых и, в том числе, экономического кадастра промышленных месторождений цветных металлов требует формирования мощной информационной базы. Эта база должна быть не только оперативно доступной (таким свойством может обладать и хорошо организованная картотека), но обеспечивать возможность решения большого развивающегося класса задач, направленных на совершенствование научного управления минерально-сырьевыми ресурсами.

Принимая во внимание весьма большое количество объектов (месторождений, рудников, комбинатов и т.п.), значительный набор показателей, определяющих эти объекты, и существенную сложность задач, мы должны констатировать, что для достижения указанных целей необходимо применение развитой информационной системы (ИС).

В настоящее время в СССР и за рубежом уже существуют и интенсивно разрабатываются многие сотни информационных систем различного целевого назначения, функциональных возможностей и технических характеристик. Для примера можно привести названия наиболее мощных ИС, существующих в СССР и рекомендуемых в промышленную эксплуатацию: БАНК, ИНЭС, ПАЛЬМА, СЕТОР, НАБОБ, СИОД, ОКА, КАМА, КОМПАС, ПАРМА, ДИЛОС, ПОЭТ, РЕФЕРАТ, АССИСТЕНТ, СИРЕНА, ПРИЗ и т.д.

Однако, ни одна из указанных выше ИС не обладает совокупностью функциональных и технических характеристик, которые позволили бы применить их без значительных доработок для создания системы управления минерально-сырьевыми ресурсами и решения большого комплекса задач на базе кадастров месторождений полезных ископаемых. Совокупность таких характеристик сводится к следующим основным возможностям, которые должна обеспечивать ИС:

I. Создание и ведение мощной разветвленной динамически изменяющейся информационной базы.

2. Непосредственное взаимодействие пользователей с ИС на профессиональном естественном языке.

3. Постановка и решение неформализуемых и плохо формализуемых задач посредством общесистемного аппарата, осуществляющего на единой программной основе процесс решения.

4. Полная независимость описания предметной области (базы знаний), определения данных (базы данных) и определения задач (базы моделей задач).

5. Определение постановок задач, локализация параметров, областей их существования, критериев и т.д. непрограммными средствами, но с применением общесистемных механизмов.

6. Организация процессов порождения информации и решения задач, распознавания запросов (постановок задач) и синтеза ответов, сохранения следов процессов решения, промежуточных состояний и расстановку контрольных точек с возможностью возврата к ним для компенсации неудач и выполнения многовариантных расчетов и т.д.

Примером информационной системы, которая развивается в направлении осуществления перечисленных возможностей и которая уже в настоящее время по многим своим характеристикам не уступает перечисленным выше ИС, является Обучаемая диалоговая информационная система с естественным языком (ОДИССЕЯ). ИС ОДИССЕЯ разрабатывается в ЛПИ и ее предметно ориентированная версия, называемая ИС «Кадастр", промышленно эксплуатируется в Главном вычислительном центре Минцветмета СССР.

Одной из важных характеристик информационных систем является быстродействие, достигаемое при выполнении процессов решения задач, которое складывается в простейшем случае из трех компонент:

I) распознавание типа задачи и ее входных параметров;

2) поиск необходимой для решения задачи информации, хранящейся в базе данных системы;

3) выполнение собственно вычислительного (и логического) алгоритма решения задачи.

Если последняя компонента сравнительно мало влияет на быстродействие ИС, то метод реализации первых двух компонент решающим образом влияет на величину быстродействия.

В этой связи представляется интересным сравнить ИС ОДИССЕЯ с системой близкого класса, т.е. с ИС, использующей в качестве языка общения с пользователем естественный язык. Такой системой является ИС ПОЭТ [7][x].

Приведем пример близких по существу запросов в сравниваемых системах, отличающихся предметными областями приложений:

ПОЭТ: Чему равен в I975 г. удельный вес перевозок железнодорожным транспортом в общем объеме перевозок всеми видами транспорта?

ОДИССЕЯ: Определить удельный вес суммарного объема добычи руды по месторождениям СССР, выбывающим из эксплуатации до I988 г., в суммарном объеме по всем эксплуатируемым месторождениям.

Время ответа системы ПОЭТ на вопрос приведенного типа около I5 с [7] при ее функционировании на ЭВМ ЕС-I060. Время ответа ИС ОДИССЕЯ при ее функционировании на ЭВМ ЕС-I022 на поставленный вопрос I40 с. Поскольку

---

[x] Заметим, что системы ПОЭТ и ОДИССЕЯ являются представителями весьма немногочисленного класса промышленно эксплуатирующихся как в СССР, так и за рубежом систем с естественным языком.

техническое быстродействие ЕС-1060 составляет I млн операций в секунду, а быстродействие ЕС-1022 – 80 тыс. операций, то ясно, что функциональное быстродействие ИС ОДИССЕЯ не уступает быстродействию ИС ПОЭТ.

Направленность ИС ОДИССЕЯ на естественный язык в качестве языка общения предопределяет очень важные свойства системы, которые значительно расширяют ее функциональные возможности. Аналогичной точки зрения на роль естественного языка общения человека с ЭВМ придерживается А.П.Ершов [5], хотя существуют и развиваются другие подходы к организации коммуникации человека с ЭВМ [4].

Концепция применения естественного языка последовательно распространяется в ИС ОДИССЕЯ на все компоненты ее функциональной организации, обеспечивая тем самым адекватность всех её структур и всех развиваемых в ней механизмов. Эффективность такого подхода многократно может быть подтверждена как потребностью практической реализации слабо формализуемых задач, так и необходимостью отображения в систему сложно организованной информационной среды. Практическая работа авторов над разработкой ИС ОДИССЕЯ в полной мере доказывает это. В ИС ОДИССЕЯ практически осуществляется тезис о необходимости взаимно однозначного соответствия языка описания информационной среды, профессионального языка пользователей и языка постановок и определений задач. Реализация этого тезиса в системе заключается в определении однозначного отображения информационной среды и среды пользователей [6] в структуры и средства системы.

Рациональность развиваемой естественноязыковой концепции хорошо иллюстрируется на примере порождаемой базы данных как однозначного образа экономических кадастров месторождений полезных ископаемых. Кадастры имеют весьма сложную иерархическую структуру (на логическом уровне) и еще более сложную лингвистическую структуру. Последняя определяется многообразием геологической, горной, экономической и технологической терминологии, а также нетривиальными отношениями, определяемыми на используемых понятиях.

Применение естественного языка для определения конкретной информации, отображаемой в базу данных, позволяет затем формулировать на том же языке задачи (запросы) и строить на его же основе методы распознавания и поиска необходимой информации из базы данных. В отличие от системы ПОЭТ, которая использует "чужеродную" базу данных как внешнюю по отношению к системе структуру со своими собственными принципами организации, методами доступа и языком описания данных, ИС ОДИССЕЯ имеет единую организацию всех структур, в частности, базы знаний, базы данных и базы моделей задач. Эта особенность ИС ОДИССЕЯ, дополненная языковым единством на уровне описания информационной среды, постановок задач и формулировок запросов, является принципиально важным свойством, обеспечивающим эволюцию функциональных возможностей системы. Определение основных принципов порождения структур системы ОДИССЕЯ и организация основных структур ее даны в работах [I, 6].

Специфический характер задач управления минерально-сырьевыми ресурсами накладывает дополнительные требования к организации ИС ОДИССЕЯ. Многие из этих задач не могут быть формально поставлены, и для них не существует

аналитических решений. Таковы, например, сложные задачи экономической оценки месторождений полезных ископаемых, поиск объекта – аналога при проектировании нового объекта (рудника, фабрики, комбината и т.п.) [2, 3]. Постановка и алгоритмическое определение таких задач является сложным процессом и требует от специалиста высокой квалификации, разработки эвристических алгоритмов, а также его непосредственного участия в процессе решения задач. Эти свойства задач приводят к необходимости развития в ИС ОДИССЕЯ специальных механизмов взаимодействия с пользователем-специалистом непосредственно в процессе решения задачи для оценки промежуточных и окончательных результатов, выбора критериев и ограничений, метода решения, определения дополнительных условий, динамического задания параметров, повторения анализа и расчета с заданной контрольной точки и т.д. Помимо того, что такие требования к системе определяют язык коммуникации, они одновременно обуславливают структурную организацию системы и средства оперирования структурами, порождаемыми в процессе решения задачи.

Использование естественного языка для описания информационной среды обеспечивает полное сохранение ее свойств и отношений. Но порождаемая при этом сложность структур и тем более последующая сложность анализа их в базе данных и в базе знаний системы вызваны прежде всего свойствами естественного языка с присущими ему синонимией, омонимией, полисемичностью и т.п., а затем уже сложностью логической организации самой информационной среды. Эти особенности языка и среды заставляют при создании ИС идти на рациональные компромиссы, которые сводятся к введению набора некоторых ограничений на язык описания информационной среды, в том числе на его синтаксис и семантику. Эти ограничения одновременно учитывают технологичность и некоторые другие эксплуатационные характеристики ИС ОДИССЕЯ.

В качестве примера приведем принятую в системе форму представления конкретной информации по объектам, показателям и значениям показателей в виде скобочных записей, структура которых отражает логическую структуру исходных данных. Эти записи не имеют жесткого формата, не требуют предварительного кодирования информации, допускают изменение возможных разделителей и т.д. Для идентификации записей различного семантического значения определяются типы записей (все данные в примерах условные):

1. Тип записи для описания объектов

Р2 (СССР ¤ I (УКРАИНСКАЯ ССР ¤ I2I (РОСТОВСКАЯ ОБЛАСТЬ ¤ 303 (ШАХТИНСКИЙ РАЙОН ¤ 4004 (РУДНИК ПЕРВОМАЙСКИЙ ¤ 5001))))) @

2. Тип записи для описания показателей

Р2 (ГОДОВАЯ ПРИБЫЛЬ(ОТ РЕАЛИЗАЦИИ ТОВАРНОЙ ПРОДУКЦИИ С УЧЕТОМ МЕТАЛЛУРГИЧЕСКОГО ПЕРЕДЕЛА ¤ I22)) @

3. Тип записи для описания значений показателей

РI9 (РI7 (I7 # 26 Р23 (РУДА (200 ТЫС.Т) СВИНЕЦ (ОТ I0000 Т ДО I2000 Т) ЦИНК (НЕ ОПРЕДЕЛЕНО))) РI7 (I7 #II # ПРЕДЛОЖЕНИЯ О ЦЕЛЕСООБРАЗНОСТИ СНЯТИЯ С БАЛАНСА НЕКОТОРЫХ МЕСТОРОЖДЕНИЙ, "КАВКАЗГИПРОЦВЕТМЕТ" I974 Г.)) @

В приведенных примерах в записи типа P2 после разделителя ¤ фиксируются номера (коды) объектов и показателей соответственно, а в записи типа PI7 эти номера используются для идентификации пар: объект – показатель (эти номера разделяются знаком # и могут быть заменены на соответствующие названия объектов и показателей).

Построенная в ИС ОДИССЕЯ грамматика входных записей для порождения информации в базе данных системы обеспечивает описание данных технологически достаточно просто и оперативно. Необходимые для описания полного кадастра (I33 показателя) по одному объекту затраты времени составляют приблизительно 3–4 ч. Функционирующая в настоящее время версия ИС ОДИССЕЯ при ее эксплуатации на ЭВМ ЕС-I022 обеспечивает загрузку одного кадастра в базу данных в среднем за 2I мин, т.е. на запись, соответствующую одному значению показателя, затрачивается в среднем I0 с.

Получаемая в результате структура базы данных сохраняет полностью логическую структуру исходной информации, а также всю присущую ей лексику, синтаксис и семантику естественного языка. В этом смысле ИС ОДИССЕЯ сталкивается со многими проблемами анализа естественного языка как на стадии распознавания запросов и формулировок задач, так и на стадии подготовки данных для их решения. Экспериментальный опыт решения задач на базе экономических кадастров месторождений цветных металлов в составе ИС ОДИССЕЯ показывает, что разработанные в ней средства достаточно успешно разрешают проблемы анализа предложений профессионального естественного языка и служат таким образом фундаментом для создания на ее основе развитой системы для решения задач управления минерально-сырьевыми ресурсами.

ЛИТЕРАТУРА

1. Арский Ю.М., Лившиц А.Н., Гатов Т.А. и др. Управление минеральными ресурсами и экономический кадастр промышленных месторождений цветных металлов. М., изд. ЦНИИЭИцветмет, 1980, вып. 4.

2. Гатов Т.А. Рациональное использование месторождений цветных металлов. М., Недра, 1980.

3. Гатов Т.А. Экономическая оценка месторождений цветных металлов. М., Недра, 1975.

4. Довгялло А.М. Диалог пользователя с ЭВМ. Основы проектирования и реализации. Киев, Наукова думка, 1981.

5. Ершов А.П. К методологии построения диалоговых систем: феномен деловой прозы. Новосибирск, изд. ВЦ СО АН СССР, 1979.

6. Лившиц А.Н., Школьников А.Д. Некоторые вопросы создания информационных систем с ограниченным естественным языком. – Труды 17-го Международного симпозиума по применению ЭВМ и математических методов в горных отраслях промышленности. М., Недра, 1982.

7. Попов Э.В. Общение с ЭВМ на естественном языке. М., Наука, 1982.