

УДК 681.3 : 639.2

ПРИНЦИПЫ ПОСТРОЕНИЯ АВТОМАТИЗИРОВАННОЙ
ИНФОРМАЦИОННО-СПРАВОЧНОЙ СИСТЕМЫ "ПРОГНОЗ"

Э.И.Ермаченко,
Э.И.Черный

Автоматизированная информационно-справочная система (ИСС) "Прогноз" предназначена для автоматизации процессов информационного обеспечения автоматизированной системы "Сырьевая база" (АССБ).

ИСС "Прогноз" относится к классу фактографических автоматизированных информационных систем (АИС) и представляет собой совокупность языковых средств, порядка хранения, организации и поиска информации и средств вычислительной техники (Глушков и др., 1973). Система выполняет следующие операции:

сбор, классификация и длительное хранение больших объемов систематизированной информации, имеющей сложную внутреннюю структуру;

ввод и обновление хранимой информации в ЭВМ и ее подготовка для последующей обработки;

формирование справочно-прогностической информации путем выполнения ассоциативного поиска в массивах накопленной информации и последующего осуществления требуемых содержательных и (или) структурных преобразований;

выдача справочно-прогностической информации в форме, пригодной для непосредственного использования человеком или последующей обработки в автоматизированной системе обработки данных.

Тип ИСС. В основу принципов построения ИСС для АССБ легли следующие отечественные и зарубежные теоретические разработки:

применение системного подхода в создании больших систем;
методы теории информатики (вопросы построения АСУ и информационных языков, методы семиотических исследований);

модельная теория мышления (методы формирования абстрактных понятий, операционального описания объектов и эвристического программирования);

современные методы машинного программирования (алгоритмические языки программирования, методы эффективной организации информации в памяти ЭВМ, принципы построения операционных систем программирования);

математико-лингвистические методы.

Разработка отдельных фактографических АИС ведется и в Советском Союзе, и за рубежом. В 1969 г. советские исследователи А.А.Стогний и Н.Г.Зайцев (1969) обобщили опыт разработок фактографических АИС и сформулировали назначение, характеристики и основные требования к ИСС.

В 1973 г. в Институте кибернетики (ИК) АН УССР под руководством А.А.Стогния и В.Н.Афанасьева был разработан технический проект типовой ИСС документально-фактографического типа. В этой работе обобщены результаты многолетних исследований и разработок, проводившихся в ИК АН УССР и выделены следующие методологические принципы построения типовой ИСС.

Система рассматривается как совокупность средств математического и организационно-технического обеспечения, позволяющих на серийно выпускаемом оборудовании эффективно реализовать автоматизацию решения информационно-справочных задач (ИСЗ).

Для пользователя, внедряющего или создающего у себя систему, она представляет собой совокупность документации: описание информационной системы и инструктивные материалы.

Совокупность средств матобеспечения ИСС может использоваться в любой предметной области с определенной привязкой к терминологии, информационным и логико-арифметическим характеристикам конкретных автоматизируемых задач и процессов.

Перечисленные общие принципы построения типовой ИСС наиболее полно удовлетворяют разработанным нами требованиям к построению ИСС "Прогноз".

Анализ специфических особенностей информационного обеспечения АССБ (Ермаченко, Черный, Опубликовано в данном томе) показал, что ИСС "Прогноз" должна быть фактографической ИСС информационно-логического типа (ИСС/ИЛТ).

По определению "Энциклопедии кибернетики" (1974), информационно-логической системой является автоматизированная система, алгоритмически решающая на основе хранящихся в ней массивов фактических данных различного рода задачи по синтезу новых сведений, не содержащихся в этих массивах в явной форме. Некоторые принципы построения ИСС/ИЛТ изложены в работе Э.И. Ермаченко (1970).

ИСС/ИЛТ относится к наиболее сложному классу фактографических АИС, так как использует формальный аппарат логико-арифметической обработки данных, не привязанный к решению конкретных задач автоматизированной системы управления. В основу создания ИСС/ИЛТ положено отделение информационно-логических процессов, таких как анализ и синтез фактов, трансформация одних фактов или понятий в другие (при небольшом количестве расчетно-вычислительных операций), от задач обработки данных. Такое выделение информационно-логических процессов в ИСС/ИЛТ позволяет этой системе обслуживать практически неограниченное количество задач, информация о которых содержится в поисковом массиве и хранится в унифицированном магнитном хранилище.

С учетом специфических особенностей технико-прогностической информации (Ермаченко, Черный, Опубликовано в данном томе) предложен новый метод организации информационной модели АССБ, позволяющий отойти от традиционных способов построения поискового массива и строить его в виде знаковых описаний содержательного состава сведений ИФ, образующих так называемую семиотическую модель ИФ. В семиотической модели устанавливаются смысловые отношения между единицами (классами объектов и описывающих их признаков) ИФ и отображается физическое расположение массивов и их составных частей в магнитных хранилищах ИФ.

Семиотическое описание информационных объектов дает возможность операционной системе "осмыслить" запрос и автоматически, без участия программиста, составить алгоритмы формирования ответа на запрос на языке ЭВМ, осуществить поиск

запрашиваемых сведений в соответствующих информационных массивах, произвести над ними логико-арифметические операции и выдать ответ в удобном для абонента виде.

Принципиальная схема ИСС. ИСС "Прогноз" ориентирована на автоматизацию таких основных процессов информационного обеспечения АССБ, как выдача технико-прогностической информации на основе ретроспективного поиска в массивах информационного фонда центра данных (ЦД) и решение фиксированного набора типовых задач по информационному обеспечению пользователей ИСС, на базе избирательного распределения технико-прогностической информации, на основе решения нестандартных задач аналитической обработки данных и документирования выдаваемой пользователю информации и на базе результатов решения специальных задач аналитической обработки данных, хранящихся в специальных массивах (в абонентском архиве) банка данных.

Соответственно принципу интеграции технико-прогностической информации, принятому в АССБ, все ЦД должны быть оснащены типовой ИСС. Такая интеграция может быть достигнута за счет информационной совместимости центров данных и требует разработки типового математического обеспечения ИСС, единой структуры типового банка данных (БД), типового математического обеспечения ведения БД, единой технологии, обеспечивающей функционирование БД, ИСС и объединяющей банки данных всех ЦД в БД АССБ и единой системы языковых средств общения с типовым БД.

В соответствии с перечисленными требованиями к информационному обеспечению АССБ разработана принципиальная схема функционирования типовой ИСС "Прогноз" (рисунок), решающей следующие задачи (номера задач соответствуют указанным в схеме):

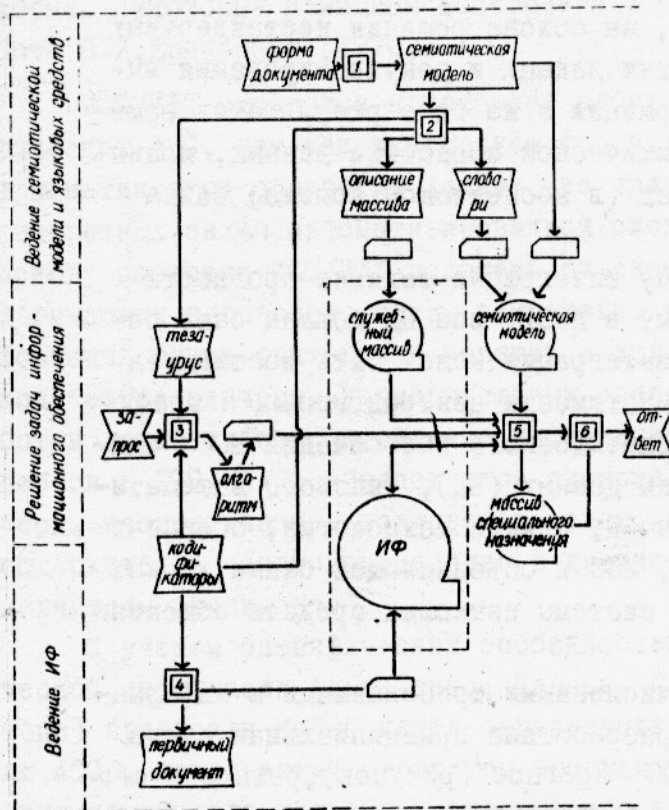
- 1) наращивание разделов семиотической модели ИФ;
- 2) запись разделов семиотической модели на магнитной ленте (МЛ): формирование описаний новых массивов и обновление служебных массивов, разработка разделов словаря для качественных признаков, перфорация разделов семиотической модели, описаний массивов, разделов словарей, запись на МЛ;
- 3) формирование текста запроса на естественном языке: индексация смысловых единиц запроса с помощью тезауруса, кодирование значений поисковых признаков при помощи кодифи-

каторов, поиск управляющей информации (УИ) и соответствующего алгоритма запроса, ввод УИ в ЭВМ;

4) запись первичной информации в ЭВМ: перфорация первичного документа, ввод первичного документа и формирование машинного документа;

5) формирование выходного массива и запись его в массивы специального назначения;

6) вывод выходного массива на алфавитно-цифровое печатающее устройство (АЦПУ) или дисплей.



Принципиальная схема работы ИСС "Прогноз"

Языковые средства ИСС. Информационный язык (ИЯ) ИСС представляет собой специализированный искусственный язык, позволяющий вести диалог в режиме "абонент-ИСС". Этот язык понятен исследователю, работающему в сфере АССБ, и пригоден для представления данных в ИСС. ИЯ строится на базе естественного языка, на котором заполняются первичные документы. В связи с этим информация на бланках первичных и вторичных документов должна достаточно полно и по возможности лаконично отражать поведение биологических объектов и состояние среды. Каждая запись (строка или зона) документа рассматривается в ИЯ как законченное предложение, а отсутствие кон-

текста требует от каждого слова или фразы ИЯ строго однозначного смысла. Это означает, что из словарного состава ИЯ должны быть исключены омонимы и синонимы, многозначные слова, неясные случаи и т.п. В то же время ИЯ должен быть приспособлен для использования его в ЭВМ. Опираясь со словами и фразами, хранящимися в памяти ЭВМ, ИСС не может обращаться к их интуитивно известному смыслу и должна ограничиваться анализом их формальной структуры.

Исходя из этого разработка ИЯ ведется нами в виде формальной семиотической системы, состоящей из следующих элементов:

- 1) лексики ИЯ;
- 2) правил образования, устанавливающих, какие комбинации лексических единиц в виде терминов, фраз или предложений допускаются;
- 3) правил преобразования, устанавливающих, какие логико-арифметические операции с элементами информационного фонда допустимы для получения ответа на информационный вопрос;
- 4) правил интерпретации, устанавливающих, какой смысл надлежит приписывать выражениям, составленным по правилам преобразования.

ИЯ разрабатывается в виде совокупности трех взаимосвязанных информационных языков: языка описания информационного фонда и его элементов (ЯИФ); языка формирования запроса (ЯФЗ) и языка формирования ответа (ЯФО).

В основу разработки ЯИФ легли наши исследования по семиотическому описанию информационных объектов (Ермаченко, Синева. Опубликовано в данном томе).

ЯФЗ разрабатывается нами в виде языка спецификации, в котором отражены некоторые принципы входного языка в информационно-справочной системе "Бейсбол" (Грин и др., 1967).

Для построения лексико-семантической основы ИЯ ИСС предусмотрена разработка тезауруса "Сырьевая база", системы классификаторов и кодификаторов и системы словарей. Тезаурус "Сырьевая база" включает классификационную схему технико-прогностических показателей АССБ, словари лексических единиц ИЯ, глоссарий и указатели.

На базе классификационной схемы технико-прогностических показателей АССБ строится семиотическая модель ИФ АССБ.

Классификаторы технико-прогностической информации представляют собой систематизированный свод классификационных

группировок значений смысловых единиц тезауруса и их кодовых обозначений. Для смысловых единиц тезауруса, значения которых не требуют классификационных группировок, целесообразно разрабатывать кодификаторы.

Словари содержат сокращенные наименования отдельных позиций классификатора, предназначенных для печати в выходных формах текста.

Организация информационной базы. При разработке ИСС были приняты два способа представления информационной модели: информационный и метаинформационный.

При информационном способе представления элементами информационной модели являются непосредственные описания информационных объектов. Совокупность этих описаний составляет информационный фонд.

При метаинформационном способе представления элементами информационной модели служат знаковые описания классов информационных объектов, причем каждый класс рассматривается как целое. Знаковое описание класса (мы назвали его концептом) осуществляется не перечислением его объектов, а установлением его отношения к другим классам и определением перечня общих для множества объектов, составляющих данный класс признаков и их соотношений.

Совокупность признаков концепта, по которым объекты обобщаются в класс, составляет содержание концепта. Достаточность и необходимость признаков, взятых в качестве содержания концепта, можно рассматривать как неперемное условие правильного образования концепта.

Множество объектов, обобщаемых данным концептом в класс, именуется объемом концепта. Отдельные объекты этого класса называются элементами класса или элементами объема концепта, а подклассы его — частями класса или частями объема концепта.

Знаковую систему концептов и их признаков, организованную на основе установленных между ними отношений и зависимостей, мы назвали семиотической моделью.

Семиотическая модель в ИСС играет роль путеводителя в поиске нужного ответа на вопрос в лабиринтах ассоциативных связей между объектами и их признаками. Семиотическая модель в ИСС выполняет следующие функции:

- 1) классифицирует первичные документы и хранящиеся в них сведения;
- 2) отражает семантику информации, хранящейся в первичных документах в такой форме, которая может быть принята и интерпретирована обслуживающим персоналом и абонентами ИСС;
- 3) классифицирует сведения, хранящиеся в информационных массивах ЭВМ;
- 4) отражает семантику информации, хранящейся в массивах ЭВМ в такой форме, которая может быть понята и интерпретирована операционной системой ИСС, работающей в режиме "запрос-ответ".

Построенная по этому принципу информационная модель позволяет вести информационный поиск по следующей схеме: вначале по некоторой совокупности признаков выделить класс (множество) информационных объектов, к которому относятся искомые сведения, затем в структурных массивах ИФ при помощи отношений отыскать адреса искомого класса сведений в фактографии ИФ и, наконец, в фактографических массивах ИФ по заданным значениям признаков собрать искомые сведения и на их основе формировать ответ.

Содержательный анализ информационной базы. Цель содержательного анализа информационной базы - построение системы технико-прогностических показателей, на основе которой строятся тезаурус "Сырьевая база" и семиотическая модель ИФ, выбирается принцип организации массивов ИФ по содержанию и классифицируются типовые запросы для первой очереди ИСС.

Содержательный анализ представляет собой последовательность семиотических исследований содержания базовой информации АССБ. Семиотические исследования ИФ ведутся двумя основными методами анализа: качественным и количественным. Качественный метод анализа сведений ИФ включает исследования, при помощи которых обнаруживаются эксплуатационные свойства ИФ (функционирование в режиме "запрос-ответ", элементы и структура, а также взаимосвязи свойств ИФ, его элементов и структуры). Основываясь на обобщении опыта разработок автоматизированных информационных систем и математико-лингвистических исследований (Ермаченко, Синева. Опубликована в данном томе), мы включили в круг исследований ка-

чественных свойств ИФ выбор макроструктуры организации сведений в ИФ, анализ лексики языка описания смыслового содержания ИФ, разработку схем потоков сообщений, образующих на базе сведений ИФ, и формирование схемы классификации сведений ИФ.

Количественный метод анализа сведений информационного комплекса позволяет доказать или отвергнуть выбранную разработчиками организацию информационного комплекса. Количественный метод в семиотических исследованиях играет эвристическую роль. Числовые выражения результатов, полученные в ходе измерений параметров информационного комплекса, позволяют дать соответствующую оценку его организации и определить направление дальнейших исследований.

На основании анализа задач, решаемых в составе первой очереди АССБ на базе сведений новых форм сбора информации, определен состав ИФ и рассчитаны возможные объемы информации для отдельных ЦД. По расчетам, годовой объем ИФ АССБ на первом этапе составит более 5 млрд.байт, а через десять лет превысит 10 млрд.байт.

В процессе исследования принципа организации базовой информации ИФ был проведен сравнительный анализ двух вариантов организации информационного хранилища ЦД.

Первый вариант хранилища предполагает организацию информации по схеме „супер-массив → массив → блок“.

Второй вариант хранилища предполагает организацию информации по схеме "супер-массив → раздел → том → массив → блок". В этой схеме "том" соответствует совокупности информации одного информационного класса за один год или несколько лет. В соответствии с этой схемой массивы хранят информацию по одной из форм первичной информации за определенный для данного информационного класса промежуток времени (совокупность лет, год, сезон, месяц), и объем массива не может превышать объем бобины МЛ.

Сравнительные характеристики основных параметров информационного комплекса (ИК) типового ИФ приведены в таблице.

Математическое обеспечение ИСС. Создание информационного фонда и регулярная его корректировка – трудоемкое и наиболее ответственное звено в цепи задач реализации ИСС, что связано с многообразием форм первичных документов, имеющих обращение в АССБ.

**Сравнительная характеристика основных параметров
информационного комплекса (ИК) типового ИФ**

Ч и с л о							Среднее время поиска и чтения, мин.					
супер-массив	разделов	томов	массивов	тис. блоков	млн. записей	Объем ИК, млн. байт	ИК	супер-массива	раздела	тома	массива	блока
1975 г.												
<u>4</u>	-	<u>32</u>	<u>27</u>	<u>4</u>	<u>34,5</u>	<u>189,5</u>	<u>II3</u>	<u>28</u>	-	<u>3,9</u>	<u>4,7</u>	<u>2,4</u>
4	27	60	189	4	34,5	189,5	II3	28	4,9	1,9	1,3	0,9
1985 г.												
<u>4</u>	-	<u>55</u>	<u>27</u>	<u>8,9</u>	<u>76</u>	<u>414</u>	<u>249</u>	<u>62</u>	-	<u>4,5</u>	<u>8,9</u>	<u>4,5</u>
4	27	66	459	8,9	76	414	249	62	9,2	3,7	2,1	1,8

Примечание. Над чертой - первый вариант, под чертой - второй.

Математическое обеспечение (МО) ИФ ЦД должно обеспечить решение задач, связанных с накоплением, корректировкой и обработкой информации в ЭВМ, в установленные сроки и с наименьшими трудовыми затратами и максимально облегчить функционирование всех служб ЦД АССБ, связанных с вводом, хранением и обработкой информационных массивов.

Унифицировать процессы ввода первичной документации в ЭВМ и преобразовать их в информационные массивы удалось благодаря разработке единой структуры организации первичных документов, информационных массивов и их описаний; единой технологии ввода, хранения и обработки унифицированной отчетно-аналитической информации и системы классификаторов и шифраторов для кодирования качественных характеристик.

Поскольку МО является составной частью операционной системы ИСС и в то же время относится к общесистемному МО АССБ, необходимо было учесть следующие положения:

1) МО должно ориентироваться на широкий круг устройств сбора, перфорации, контроля и хранения, которые непрерывно изменяются качественно и количественно.

2) успех функционирования МО обеспечивается единством технологического процесса ведения информационных массивов,

связывающего в единый коллектив специалистов широкого профиля;

3) МО должно учитывать изменения в структуре, содержании и средствах контроля информационных массивов в связи с непрерывным развитием функций и задач прогнозирования.

С учетом этих положений к МО были сформулированы следующие требования (Ермаченко, 1972):

1) стандартные программы МО должны разрабатываться в виде типовых задач ИСС;

2) алгоритмы стандартных программ должны быть дифференцированы на модули, чтобы в случае возникновения различного рода изменений в технологии ведения массивов можно было бы оперативно заменить небольшое число модулей;

3) проектирование, программирование и отладка задач МО должны базироваться на единой методике.

При обработке математического обеспечения ИСС были использованы принципы модульного программирования (Ермаченко, 1972а), применяемого в настоящее время при разработке отечественных и зарубежных операционных систем.

Основное применение принцип модульного программирования нашел при разработке алгоритмов запросов. В алгоритме запроса каждый модуль не связан непосредственно с работой других модулей и предназначен выполнять одну из элементарных операций в общем процессе поиска и обработки информации. Каждый модуль связан только с информационной средой, которую образуют модульные поля. На основании сведений модульных полей модули управляют информационными процессами в рабочих полях и связывают с внешними устройствами ввода, хранения и вывода информации. При этом модульные поля выполняют функции индикаторов, которые отражают информационные процессы в массивах в любой момент времени.

Технико-организационное обеспечение АССБ. ИСС базируется на технических средствах АСУ с учетом применения дополнительных технических средств, предназначенных для обслуживания абонентов ИСС. Сюда относятся различные специальные картотеки, диктофоны, устройства для подготовки запросов, микрофильмирования, чтения и просмотра микрофильмов и т.д.

Организационно-технологическое обеспечение ИСС включает технологическую схему процессов функционирования АСУ, а также комплекс различных мероприятий, необходимых для распределения работ между персоналом ИСС по специальности и уровням компетенции.

Л и т е р а т у р а

Глушков В.М., Стогий А.А., Афанасьев В.Н. Автоматизированные информационные системы (Основные принципы построения и элементы проектирования). М., "Знание", 1973, 103 с.

Грин Б., Вулф А., Чемский К., Лоферн К. "Бейсбол" - программа, автоматически отвечающая на вопросы. - "Вычислительные машины и мышление". М., Мир, 1967, с.221-233.

Ермаченко Э.И. О некоторых принципах построения информационно-справочной системы. - "Труды 2-й Всесоюзной конференции по программированию, заседание М. Новосибирск, 1970, с.15-26.

Ермаченко Э.И. О принципах формирования информационного фонда в автоматизированной информационно-справочной системе. Материалы семинара "Информационное обеспечение АСУ", М., 1972, с.141-147.

Ермаченко Э.И. Использование принципов модульного программирования при разработке математического обеспечения автоматизированной информационно-справочной системы. Материалы семинара "Математическое обеспечение АСУ", М., 1972а, с.98-101.

Ермаченко Э.И., Синева Т.В. Некоторые принципы построения языка описания информационного фонда АССБ. Опубликовано в данном томе, с. 158-164.

Ермаченко Э.И., Черный Э.И. Некоторые вопросы построения информационного обеспечения АССБ. Опубликовано в данном томе, с.133-142.

С т о г н и й А.А., З а й ц е в Н.Г. Автоматизированные информационно-справочные системы, их назначение, характеристики и основные требования к ним. - "Кибернетика". 1969, № 4, с.124-128.

Э н ц и к л о п е д и я кибернетики, т.1, Киев, 1974, с.398-399.

Principles of construction of automated
information-reference system PROGNOZ

E.I.Ermachenko, E.I.Cherny

S u m m a r y

The automated information-reference system PROGNOZ is oriented to automation of the processes of information software in the automated system "Marine resources". The information-reference system is worked out as a retrieval system of an information-logical type. The principle components of the system are as follows: information stock, semiotical model of the information stock, language means of communication with the information stock, mathematical and technical-arranging software.