

УДК 681.3 : 639.2

НЕКОТОРЫЕ ПРИНЦИПЫ ПОСТРОЕНИЯ ЯЗЫКА  
ОПИСАНИЯ ИНФОРМАЦИОННОГО ФОНДА АССБ

Э.И.Ермаченко, Т.В.Синева

Разрабатываемая в рамках АССБ информационно-справочная система (ИСС) "Прогноз" призвана управлять информационным фондом АССБ и обеспечивать автоматическое формирование и выдачу ответов на различные вопросы потребителя.

Основная цель создания ИСС - улучшение обслуживания подсистем прогнозирования АССБ путем снабжения абонентов ИСС необходимой информацией, объем и состав которой обеспечивает своевременное формирование достоверного прогноза ("Методические основы ...", 1972).

ИСС является системой фактографического типа, т.е. она должна представлять информацию в ответ на запросы потребителя в виде конкретных фактов. Автоматическое извлечение фактографических данных из документов, хранящихся в памяти ЭВМ, - задача сложная (Штейн, 1972), поскольку вычислительная машина имеет существенный недостаток: она "не понимает" языка абонента, ей неизвестна семантика запрашиваемых сведений.

Для того чтобы ЭВМ могла успешно справиться с задачей поиска фактов в системе документов, ее память необходимо наделить некоторой моделью информационного фонда, отражающей его смысловое содержание при помощи системы взаимосвязанных сведений.

подавляющая часть данных хранится и обрабатывается в форме знаков - алфавитных и цифровых символов, поэтому проблематика исследований информации может быть описана достаточно полно в терминах теории знаковых систем - семиотики. В данной ИСС для использования в качестве информационной моде-

ли, отражающей смысловое содержание исследуемого объекта, разработана семиотическая модель (Ермаченко, Черный. Опубликовано в данном томе).

Построению семиотической модели предшествует содержательный анализ информации, циркулирующей в АССБ, который включает анализ содержания информационного фонда, входных форм сбора промыслово-биологической информации и возможных запросов к ИС.

Анализ содержания информационного фонда и анализ форм сбора позволяют выявить объем и состав информации, подлежащей хранению в АССБ. На основании этого анализа определяется макроструктура организации сведений, под которой мы понимаем абстрактную схему хранилища сведений. Единство макроструктуры организации сведений в ИФ позволяет формализовать процессы формирования запросов, поиска запрашиваемых сведений и редактирование ответа.

Выбор абстрактной схемы хранилища сведений диктовался следующими установками.

1. Макроструктура организации сведений должна отражать семантическую взаимосвязь между сведениями ИФ при минимальной затрате времени на поиск нужных сведений. В связи с этой установкой все сведения ИФ распределены по концептам (классам) и в зависимости от глубины их детализации располагаются по трем информационным ярусам: верхним, средним и нижним.

Верхние информационные ярусы ("район промысла", "информационный класс", "совокупность лет", "год" и т.д.) сами по себе не содержат фактической информации, но они дают возможность автоматически формировать различного рода сводные показатели.

Средние информационные ярусы ("шифр судна", "номер пробы", "гидрометеорологический параметр" и т.д.) отражают основные источники информации, на базе которых формируется ИФ.

Нижние информационные ярусы позволяют получать подробные сведения о состоянии сырьевой базы и формировать показатели со значительной степенью детализации.

2. Макроструктура организации сведений должна соответствовать физическим уровням организации сведений в системе документооборота АССБ. В связи с этой установкой определены основные уровни детализации (ИФ, информационный комплекс, концепт, форма первичного документа, фраза, признак), что

позволяет унифицировать организацию сведений в документальных и магнитных хранилищах ИФ.

3. Макроструктура организации сведений должна обеспечить связь ИФ с потребителями информации. Определены основные потребители информации ИФ АССБ: АСУ флота на промысле (прогнозы промысловой обстановки); центральный аппарат Минрыбхоза (справочно-прогностическая информация по широкому кругу вопросов); организации Главрыбвода (сведения для регулирования рыболовства); проектные организации (информация для проектирования орудий лова и т.д.); бассейновые институты; промысловые разведки; отдельные исследователи.

Анализ возможных запросов связан с необходимостью учета конкретных требований потребителя к ИСС. Поскольку все конкретные запросы потребителя к ИСС перечислить или сформулировать практически невозможно, такой анализ может быть сделан на основе определения типов запросов в зависимости от сложности и алгоритмов обработки информации при поиске ответа.

Анализ возможных запросов, помимо уточнения состава информационного фонда, позволяет определить, насколько рациональна организация информации в ИФ АССБ, т.е. насколько быстро, полно и точно может быть получен ответ на запрос.

Кроме того, анализ возможных запросов способствует определению стратегии и алгоритмов поиска информации и позволяет уточнить этапность разработки системы (ИСС) в зависимости от сложности обработки информации и возможностей системы на каждом этапе.

Содержательный анализ дает возможность построить семиотическую модель информационного фонда ИСС. Семиотическая модель хранится в памяти ЭВМ вне массивов информационного фонда. Она фиксирует отношения между символами в семантическом и прагматическом аспектах.

В семантическом аспекте семиотическая модель отображает смысловое содержание информации в виде формальной структуры выражений информационного языка.

Прагматические связи — это связи между информационным фондом и абонентами системы, которые отражают требования получателя к информационной системе, ценность получаемой информации.



Семиотическая модель имеет вертикальную и горизонтальные структуры. Вертикальная структура представляет собой систему связей между концептами, являющимися знаковыми описаниями классов объектов ( Ермаченко, Черный. Опубликовано в данном томе ). В вертикальной структуре все объекты классифицированы и соотносятся таким образом, что каждому объекту в информационной модели соответствует строго фиксированное место.

Горизонтальную структуру образуют горизонтальные отношения, т.е. логические отношения между признаками и их элементами в пределах одного концепта. Горизонтальная структура позволяет получать более узкие группировки из объектов класса, обобщаемого данным концептом.

В горизонтальных структурах устанавливаются два вида связей между признаками.

Первый вид связей проявляется в объединении терминов, встречающихся в одном запросе, во фразы. Между терминами во фразе нет формально-определенных отношений. Фразы фиксируют совместное появление терминов в запросе и смысловые отношения между ними. Фразы являются пересекающимися группами. Одни и те же термины могут входить в разные фразы, фразы отражают прагматические отношения и способствуют более рациональной организации информации в ИФ АССБ.

Второй вид связей в горизонтальных структурах имеет формальное выражение. Это логические отношения между терминами во фразе, которые способствуют решению задач логического и смыслового анализа информации при выдаче ответа на запрос. Эти связи представлены в виде базовых отношений. В данной системе используется пять видов таких отношений:

$r_1$  - "род - вид",  $r_2$  - "иметь в составе",  $r_3$  - "определяемое - определяющее",  $r_4$  - "целое - часть",  $r_5$  - "быть суммой".

Эти средства позволяют определять значение сложного знака (концепта) по значению более простых знаков, входящих в него. При помощи базовых отношений раскрывается содержание запроса, в котором запрашиваемые данные требуют в рамках системы дальнейшего смыслового расчленения.

При построении горизонтальных структур каждый признак -

термин — делится на обороты. Оборот — это слово или словосочетание естественного языка, которое является элементарной смысловой единицей термина.

Обороты связываются перечисленными выше отношениями в цепочку вида  $X_1 r_1 X_2 r_2 X_3 r_3$ ,

где  $X_i$  — обороты;

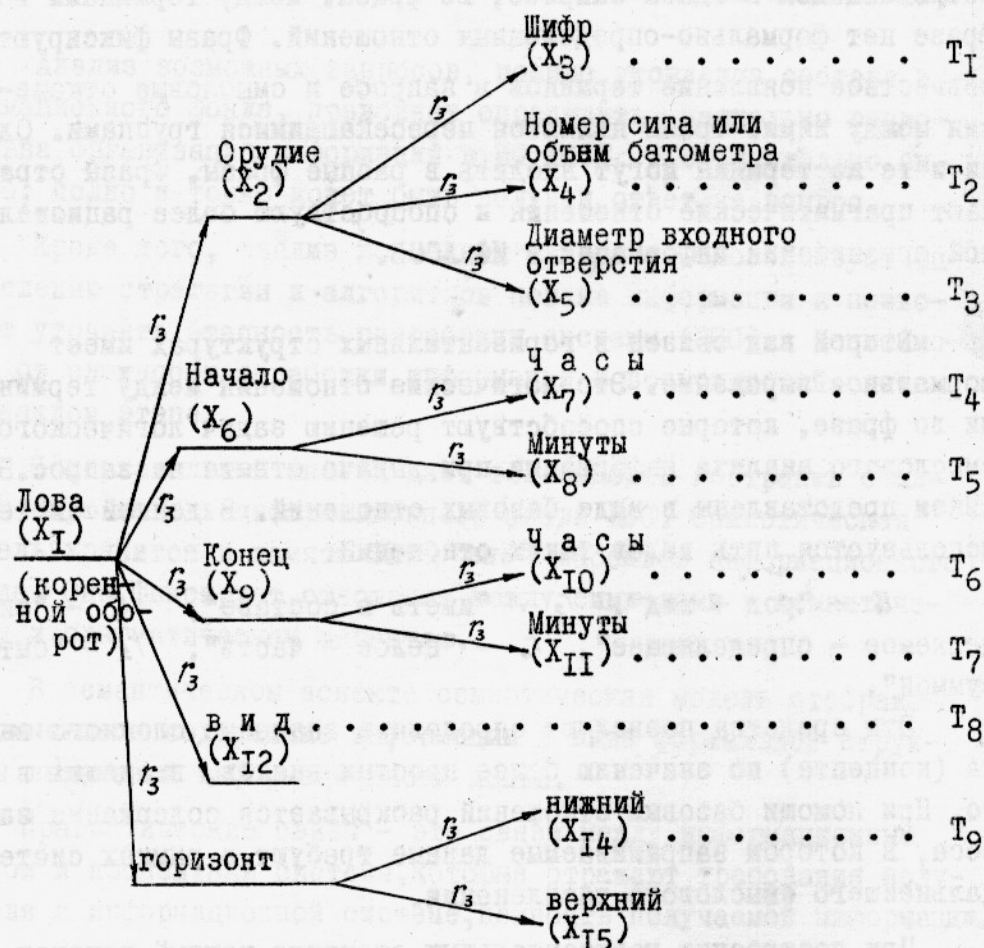
$r_i$  — базовые отношения.

Цепочка связанных между собой оборотов представляет собой структуру термина.

Основу такой цепочки составляет коренной оборот, который несет наибольшую смысловую нагрузку в данном термине, а остальные обороты в какой-то степени являются его характеристикой. Коренной оборот помещается в начало цепочки.

Эту цепочку можно представить структурно на примере горизонтальной структуры понятия "лов" (схема I).

С х е м а I



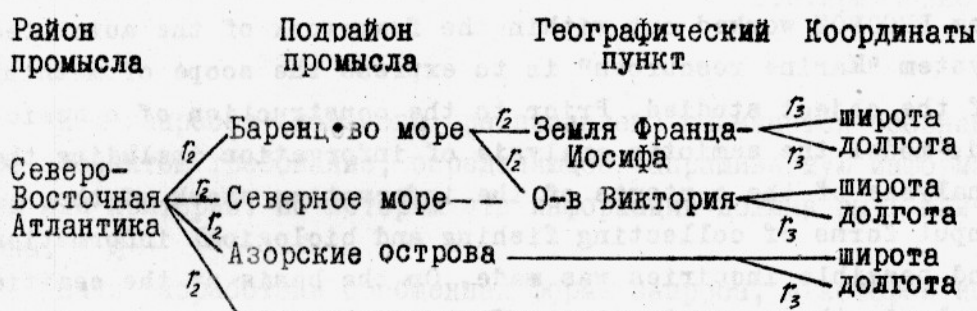
Тогда термин  $T_I$  (шифр орудия лова) можно представить в виде  $X_1 r_3 X_2 r_3 X_3$ ,

а термин  $T_{IO}$  (верхний горизонт лова) - в виде

$X_1 r_3 X_{13} r_3 X_{15}$

Такой сложный термин, связанный при помощи отношения "входить в состав" ( $r_2$ ) с каким-либо концептом, позволяет автоматически определить принадлежность конкретной информации к соответствующему информационному классу или комплексу. В качестве примера приведем фрагмент структуры информационных комплексов (схема 2).

С х е м а 2



Построенные таким образом горизонтальные структуры нельзя считать окончательными. Они будут уточняться по мере разработки и построения семиотической модели и корректироваться в процессе функционирования системы.

Представленный здесь язык описания информационного фонда должен явиться основой полной автоматизации обработки информации в АССБ в режиме "запрос-ответ".

### Л и т е р а т у р а

Ермаченко Э.И., Черный Э.И. Принципы построения автоматизированной информационно-справочной системы "Прогноз". Опубликовано в данном томе, с.120-132.

Методические основы разработки автоматизированной системы "Сырьевая база". М., изд.ВНИРО, 1972, 39 с.

Штейн В.С. Фактографические информационно-поисковые системы; их особенности и проблемы. НТИ сер.2, 1972, № 3, с.3-10.

Some principles of construction of the descriptive language of the information stock in the automated system "Marine resources"

E.I.Ermachenko, T.V.Sineva

S u m m a r y

The information model of the information-reference system PROGNOZ worked out within the framework of the automated system "Marine resources" is to express the scope of meaning of the object studied. Prior to the construction of a semiotic model the semiotic analysis of information including the analysis of the contents of the information stock of the input forms of collecting fishing and biological information and possible inquiries was made. On the basis of the semiotic analysis the macrostructure of arranging information allowing for formatization of the processes of formation of inquiries, search of information asked and answer output is determined. The semiotic model contains the meaning of information as a formal structure of expressions of the information language. The language means of the description of the information stock are a basis for complete automatization of processing information in the "inquiry-answer" regime of the automated system "Marine resources".