

УДК 681.3 : 639.2

О ФОРМУЛИРОВАНИИ УНИВЕРСАЛЬНОГО ЗАПРОСА
В АВТОМАТИЗИРОВАННОЙ ИНФОРМАЦИОННО-СПРАВОЧНОЙ
СИСТЕМЕ "ПРОГНОЗ"

Э.И.Ермаченко

Под запросом (поискомв предписанием) понимается осознанное абонентом требование, определяющее запрашиваемую информацию или критерии, по которым эта информация должна быть найдена.

Нами разработана обобщенная форма запроса, в которой использованы некоторые принципы языка информационно-справочной системы "Бейсбол" (Грин и др., 1967).

В качестве канонического выражения смысла запроса в ИСС в нашей интерпретации принято лингвистическое выражение запроса в виде спецификационного списка. Запрос, сформулированный таким образом, представляет собой последовательность заданий объектов запроса и описаний логико-арифметической обработки его сведений, в результате которой формируется ответ.

Каждое задание состоит из четырех частей: плана запроса, объекта запроса, критерия запроса и комментария.

План запроса определяется номером этапа запроса (в виде порядкового номера задания в запросе) и самим заданием (в виде шифра стандартной процедуры, реализующей такие стандартные процессы, как "выбрать", "подсчитать" и т.п.).

Объект запроса определяется шифром концепта (класса объектов, к которому относится в рассматриваемом задании информационный объект) и шифром термина, отражающего аспект поиска.

Критерий запроса определяет условия или ограничения значений искомых признаков. Критерий запроса описывается двумя

словам, из которых первое – оператор условия в виде одного из символов ("+", "-", "x", ":", ">", "≥", "=", "<", "≤"), а второе – операнд условия в виде числового значения характеристики объекта запроса или специального символа ("?" – искомая характеристика, "Г" – перейти к следующему оператору, "*" – запомнить значение).

Комментарий представляет собой указание абонента на место найденного или рассчитанного показателя в ответе.

Под объектом запроса мы понимаем комплекс характеристик, которые позволяют определить, каким массивам или комплектам документов соответствуют интересующие абонента сведения.

Объект запроса задается в виде последовательности заданных элементов объекта запроса:

⟨объект запроса⟩ ::= ⟨задание объекта⟩ | ⟨объект запроса⟩ ⟨задание объекта⟩

⟨задание объекта⟩ ::= ⟨индекс концепта⟩ , ⟨список признаков⟩

⟨список признаков⟩ ::= ⟨индекс признака⟩ | ⟨индекс признака⟩ ⟨список признаков⟩.

Зная объект запроса, операционная система ИСС может настроиться на просмотр информационных массивов, соответствующих заданным концептам, и обработку признаков, заданных в списках признаков.

Способ задания объекта запроса характеризуется объемом запрашиваемых сведений, глубиной их детализации, типом поисковых признаков и способом их сочетания.

В зависимости от объема запрашиваемых сведений объекты запроса можно разбить на две группы. Первую группу составляют объекты запроса единичного характера: на их базе запрашиваются сведения об отдельных конкретных объектах. Ко второй группе относятся объекты запроса сводного характера: на их базе запрашиваются сведения, обобщаемые по различным административным уровням.

Глубина детализации сведений является показателем подробности описания объекта запроса и характеризуется номером информационного яруса, к которому принадлежит самый конкретный элемент объекта запроса. В информационном фонде ИСС фактическая информация хранится на средних и нижних ярусах. Показатель глубины детализации сведений об объекте запроса прини-

мает значение "средние ярусы" или "нижние ярусы" в зависимости от того, к какому уровню относится самый конкретный элемент объекта запроса.

Все признаки, характеризующие объект запроса, мы разделили на расстановочные и нерасстановочные.

Расстановочным признаком является признак, в соответствии с конкретным значением которого расставляются документы на стеллажах или располагаются отдельные сведения об объекте на магнитных носителях. Значениями расстановочных признаков являются наименования конкретных объектов, объединяемых концептом задания объекта запроса. Расстановочные признаки при условии соблюдения принципов расположения документов или отдельных сведений в хранилищах ИСС позволяют оперативно отыскать эти документы или сведения.

Нерасстановочными признаками являются признаки, не влияющие на физический порядок расположения документов или сведений в хранилищах.

По сочетанию поисковых признаков объекты запроса подразделяются на две группы. В первую группу входят объекты запроса, поисковые признаки которых относятся к одной фразе. В данном случае поиск сведений ведется в одном аспекте. Вторую группу составляют объекты запроса, признаки которых относятся к разным фразам. В этом случае может вестись поиск в нескольких аспектах.

Целесообразно различать два уровня задания объекта запроса: абстрактный и конкретный. На абстрактном уровне в качестве объекта запроса задаются концепты и списки признаков, которые определяют множества сведений, формирующих ответ; на конкретном - значения расстановочных признаков, определяющие степень обобщения сведений, заданных на абстрактном уровне. Таким образом, одному объекту запроса, заданному на абстрактном уровне, будет соответствовать некоторое число вариантов его заданий на конкретном уровне.

Для изложения методических требований к формулированию универсального запроса потребовалось ввести ряд новых понятий: тематика информационного фонда (ИФ); алгоритм запроса, типовой запрос.

Под темой ИФ мы понимаем один или несколько аспектов смыслового содержания ИФ, зафиксированных в виде сочетания

признаков (ξ). В ИСС каждое устойчивое сочетание признаков, отражающее один из аспектов технико-прогностической информации, объединено в одну фразу. В связи с этим содержание темы может быть выражено в виде одной или нескольких фраз (Ермаченко, Синева. Опубликовано в данном томе).

Совокупность всех тем, образованных на базе фраз, описывающих концепт, составит тематику концепта. Совокупность всех тем, образованных на базе фраз, описывающих ИФ, составит тематику ИФ.

Запрос, смысловое содержание которого соответствует одной теме, назовем тематическим запросом.

Каждый из признаков, образующих содержание темы, имеет свой объем номенклатуры возможных значений. В этой связи тема может иметь некоторое количество вариантов, соответствующих возможным сочетаниям значений признаков, образующих содержание темы. Каждому варианту темы будет соответствовать один или несколько вариантов тематического запроса (в зависимости от количества типов алгоритмов, в состав которых может войти данная тема).

Алгоритм запроса представляет собой последовательность стандартных процедур поиска, реализующих формирование ответа на запрос; процедура поиска – последовательность элементарных операций с некоторым количеством сведений или их описаний. Одна и та же процедура поиска может быть реализована в виде различных сочетаний из элементарных операций. Стандартная процедура поиска представляет собой наиболее удачный вариант сочетаний элементарных операций, обеспечивающий реализацию данной процедуры поиска за наиболее короткий промежуток времени. Каждой стандартной процедуре поиска в ИСС соответствуют свои наименования и шифр.

Все стандартные процедуры поиска разбиты на две основные группы: просмотрные и логико-арифметические.

Просмотровая процедура поиска заключается в переборе сведений о множестве объектов в четырех аспектах: по способу просмотра, количеству ее ступеней, способу хранения сведений и направлению поиска.

По способу просмотра стандартные процедуры поиска подразделяются на процедуры одноаспектного и многоаспектного

характера. Первые сводятся к просмотру сведений в зависимости от значений, принимаемых одним признаком, вторые — к перебору сведений в зависимости от значений, принимаемых несколькими признаками.

По количеству ступеней просмотра в одном аспекте стандартные процедуры поиска подразделяются на одноступенчатые и многоступенчатые. К одноступенчатым процедурам поиска относится просмотр сведений в пределах одного концепта, к многоступенчатым — просмотр в виде последовательности ступеней перебора сведений, относящихся к разным концептам одной иерархической ветви классификации.

По способу хранения сведений одноступенчатые стандартные процедуры поиска делятся на процедуры поиска в одном и в нескольких массивах.

Стандартные процедуры поиска могут вестись в прямом и обратном направлении, кроме того, поиск может носить эвристический характер.

Прямой поиск — это нахождение адреса сведения по индексу (коду) объекта, обратный — поиск индексов объектов, удовлетворяющих определенным значениям признаков запроса. Эвристический поиск включает элементы прямого и обратного поиска и применяется при запросах, требующих многоаспектного просмотра массивов. Стратегия эвристического поиска формируется в процессе самого поиска.

Логико-арифметические процедуры поиска сводятся к логико-арифметическим операциям с характеристиками, найденными на отдельных этапах просмотра: подсчет количества объектов, выбор значений признаков и расчет количественных характеристик.

Типовой запрос характеризуется способом задания объекта запроса, фиксированным составом стандартных процедур и последовательностью их выполнения в алгоритме запроса.

Универсальный запрос оформляется в виде спецификационного списка. Нами разработана обобщенная форма бланка запроса. Сформулированный на естественном языке запрос вносится абонентом в верхнюю часть бланка. Оператор формализует запрос и разносит его по графам в нижней части бланка.

Бланк запроса

Запрос на естественном языке

Найти результаты измерений сельди (I4) размером менее 20 см, пойманной в Беринговом (I5) море в феврале (2) 1963 г. на СРТ "Бирокан" (27) при температуре воды 40С, указать координаты вылова и количество измерений

1. Выбрать (3), район промысла (I6), шифр (I0I) = I5
координаты (I04) = ?

2. Выбрать (3), дата (4), год (29) = 1963
месяц (3I) = 2
число (32) = ?

3. Выбрать (3), судно (I4), шифр (75) = 27

4. Выбрать (3), воды (I9), температуру (40) = 40

5. Выбрать (3), вид рыбы (37), шифр (44) = I4
длина (90) = 20, количество (I7) = ?

| Индекс абонента | Индекс запроса | Спецификационный список | | | | | | |
|-----------------|----------------|-------------------------|---------|----------------|--------|-----------------|---------|-------------|
| | | план запроса | | объект запроса | | условие запроса | | комментарий |
| | | этап запроса | задание | концепт | термин | оператор | операнд | |
| 2I | 3 | I | 3 | I6 | I0I | = | I5 | I1 |
| | | | | | I04 | = | ? | 22 |
| | | | | | 29 | = | 1963 | I2 |
| | | 2 | 3 | 4 | 3I | = | 2 | I3 |
| | | | | | 32 | = | ? | 2I |
| | | 3 | 3 | I4 | 75 | = | 27 | I5 |
| | | 4 | 3 | I9 | 40 | = | 4 | I4 |
| | | 5 | 3 | 37 | 44 | = | I4 | I6 |
| | | | | | 90 | = | 20 | I7 |
| | | | | | I7 | = | ? | 23 |

Бланк ответа представляет собой обобщенную форму ответа на универсальный запрос. Мы предлагаем следующий порядок выдачи абоненту ответа: в начале распечатываются помеченные в графе "комментарий" наименования объектов запросов и их

значения, а затем каждый искомый показатель снабжается списком найденных или вычисленных значений. Например:

Бланк ответа

Абонент = 21, запрос = 3

Район лова = Берингово море, год = 1963, месяц = сентябрь,
температура воды = 4⁰, судно = СРТ "Бирокан",
вид рыбы = сельдь, длина рыбы = 20 см

| Число | Координаты | Количество, шт. |
|-------|--|-----------------|
| 1 | 56 ⁰ 18 - 165 ⁰ 31 | = 150 |
| 8 | 56 ⁰ 27 - 165 ⁰ 36 | = 179 |
| 17 | 56 ⁰ 28 - 165 ⁰ 40 | = 137 |

Потоки справочно-прогностических сообщений невозможно изобразить в виде традиционной схемы из-за отсутствия заранее известного смыслового состава сообщений, их адресатов и сроков получения. Учитывая это, мы поставили задачу построить классификационные таблицы, которые отражали бы достаточно полно и наглядно основные типы запросов к ИСС, модификации этих типов и их объемы.

Для решения подобной задачи информационная модель сырьевой базы рыболовства и составляющие ее классы элементов (районы промысла, биологические объекты, орудия лова и т.п.) рассматриваются нами как отдельные объекты запроса, а конкретные объекты этих классов или их совокупность — как отдельные варианты задания объектов запроса.

Зная объемы номенклатур значений отдельных признаков смыслового содержания ИФ, можно выделить основные варианты его тем и определить полный объем вариантов темы как отдельного концепта, так и ИФ в целом.

При такой трактовке содержания информационного комплекса (рисунок) объем потоков справочных сообщений можно вычислить как полную совокупность конкретных запросов к ИФ по формуле

$$N = \sum_{t=1}^T N_t$$

где t — текущий номер темы;
 T — количество тем в ИФ;

N – количество всех конкретных запросов к ИФ;
 N_t – количество конкретных запросов типа "t";

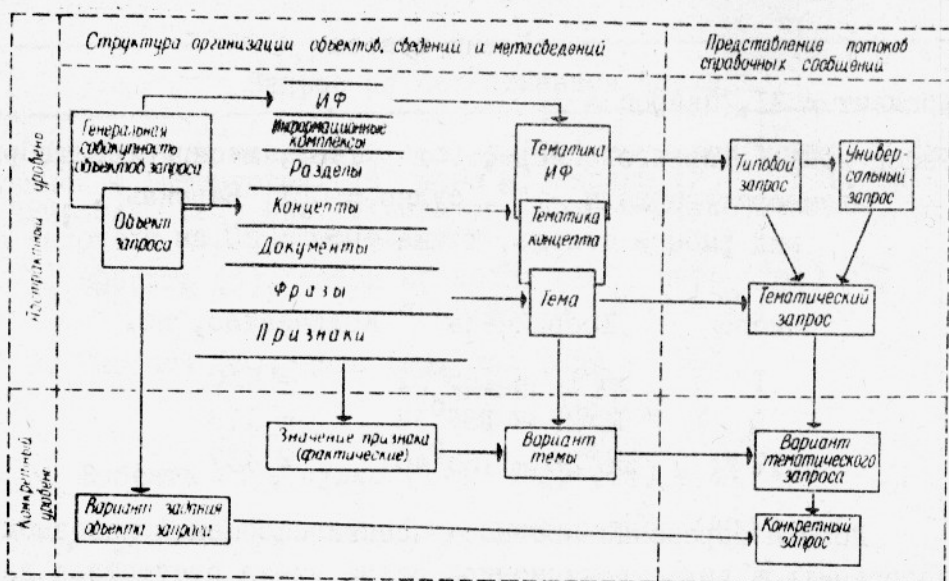


Схема организации потоков справочных сообщений в ИСС

Для объектов запросов сводного характера $\sum S'_k F'_{k,t}$,
а для объектов запросов конкретного характера $\sum S''_k F_{k,t}$,
где S'_k, S''_k – количество вариантов задания объектов запроса на базе концепта k ; $F_{k,t}$ – количество вариантов тематических запросов к концепту с алгоритмом запроса типа "t"; H, K – минимальный и максимальный номера концептов, относящихся к теме t .

Л и т е р а т у р а

- Грин Б., Вулф А., Чемский К., Лоферри К. "Бейсбол" – программа, автоматически отвечающая на вопросы. – "Вычислительные машины и мышление". М., Мир, 1967, с.221–233.
- Ермаченко Э.И., Синева Т.В. Некоторые принципы построения языка описания информационного фонда АССБ. Опубликована в данном томе, с. 158–164.
- Ермаченко Э.И., Черный Э.И. Принципы построения автоматизированной информационно-справочной системы "Прогноз". Опубликована в данном томе, с.120–132.

On formulation of universal inquiry in the
automated information-reference system
PROGNOZ

E.I.Ermachenko

S u m m a r y

As a canonical expression of the meaning of the inquiry a linguistic expression in the form of specification list is accepted. The assignment of an inquire subject is made on two levels: abstract and concrete. The following notions are introduced: a summarized form of inquiry, subjects of the information stock, algorithm of inquiry, typical and universal inquiry. The chart of arranging flows of reference information reflects the structure of arranging subjects information and meta-information on the abstract and concrete levels of presentation in the information stock.