

СООБЩЕНИЯ  
ОБЪЕДИНЕННОГО  
ИНСТИТУТА  
ЯДЕРНЫХ  
ИССЛЕДОВАНИЙ

ДУБНА



14/IV-75

4840  
A-84

P10 - 8621

Д.Д. Арнаудов

1454/2-75

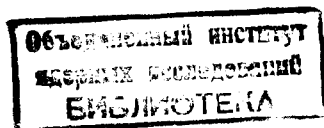
СТРУКТУРНО-ФУНКЦИОНАЛЬНАЯ ОРГАНИЗАЦИЯ  
ОСНОВНЫХ ИНФОРМАЦИОННЫХ  
МАССИВОВ ИПС ОИЯИ

**1975**

P10 - 8621

Д.Д. Арнаудов

СТРУКТУРНО-ФУНКЦИОНАЛЬНАЯ ОРГАНИЗАЦИЯ  
ОСНОВНЫХ ИНФОРМАЦИОННЫХ  
МАССИВОВ ИПС ОИЯИ



Основная проблема, которую необходимо решить при создании структурно-функциональной организации основных массивов поисковой системы - это проблема организации массива, которая фактически появляется тогда, когда поисковый массив становится настолько большим, что невозможно, или, по крайней мере, неэкономично просматривать каждый элемент массива для проверки его соответствия. Следовательно, необходимо, чтобы поисковый массив имел структуру, обеспечивающую возможность использования менее сложных механизмов индексации и промежуточных мер для определения степени соответствия, чем окончательная мера соответствия. Именно этот аспект порождает технические трудности при разработке системы информационного поиска, связанные с согласованием объема памяти, поискового массива, времени, необходимого для получения ответа, избирательной способности и точности ответа. Поэтому особое внимание должно быть обращено на структурно-функциональную организацию основных поисковых массивов, поскольку именно рациональная организация даст возможность эффективно построить систему.

Основные принципы построения эффективной ИПС при наличии большого объема информационного массива и дескрипторного словаря описаны в работе /1/.

В соответствии с указанными принципами предлагаем следующую структурно-функциональную организацию ИПС. Общая схема показана на рис.1.

В системе используются четыре типа массивов:

1. МД - массив библиографических данных о документах, называемый кратко массивом документов.

2. ОМПОД - основной массив поисковых образов документов, построенный в виде ассоциативных узловых списков.

3. МЗД - массив заголовков дескрипторов.

4. ОМД - основной массив дескрипторов.

Поясним сущность каждого из массивов.

#### Массив документов (МД)

Этот массив в поиске не участвует. К нему обращаться ние после того, как поисковые образы релевантных документов найдены.

Массив документов располагается на магнитном диске и строится как массив с прямым доступом. В общем случае он может занимать больше чем один магнитный диск.

Массив документов состоит из логических записей переменной длины. Каждая запись имеет индекс (номер записи), который в явном виде в массиве документов не указывается. Этот номер указан в ОМПОД и является ключом для прямого доступа к библиографии документа.

Библиография каждого документа является библиографической записью массива и состоит из двух частей: фиксированная часть записи и переменная часть записи. Эти записи идут одна за другой в хронологическом порядке и плотно записываются на МД.

В структурном отношении МД представляет следующую логическую структуру (см. рис.2).

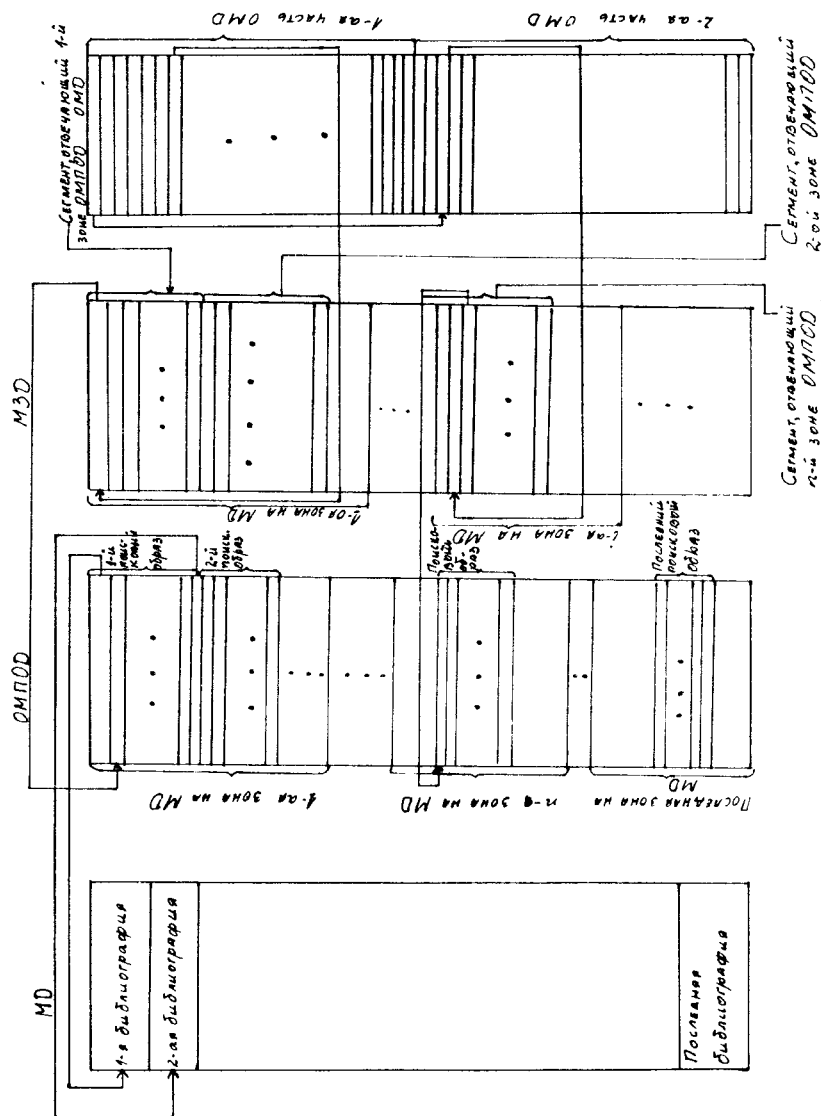


Рис.1

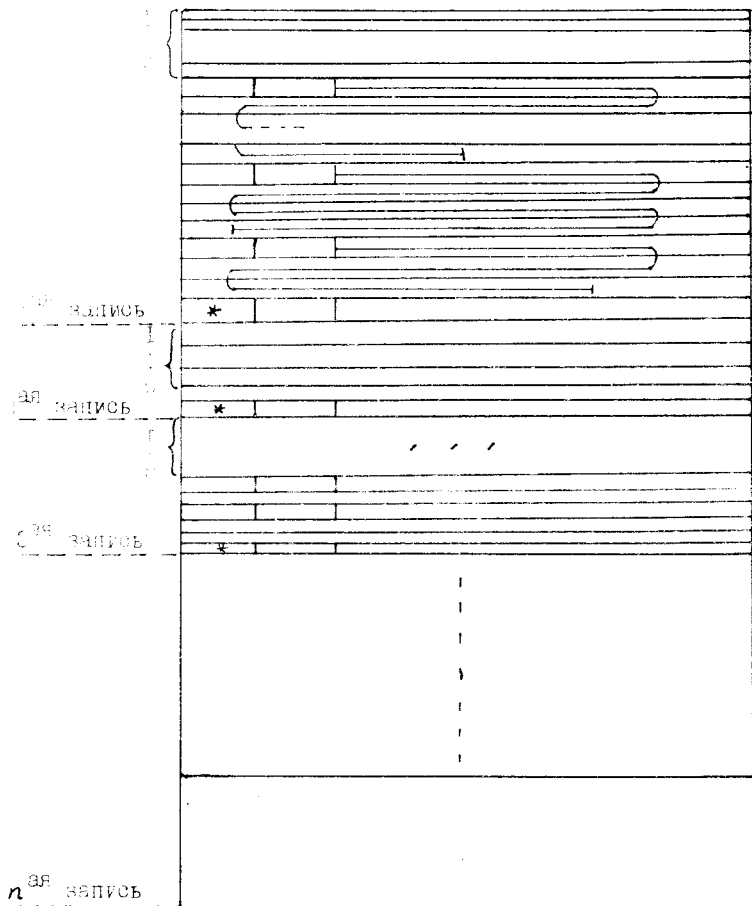


Рис.2.

Фиксированная часть библиографической записи представляет набор фиксированных полей и занимает память из 9 слов.

Начиная со следующего элемента, располагается переменная часть записи. Она заканчивается знаком (\*), который находится в начале элемента, следующего за последним элементом, где имеется информация о данной библиографической записи.

Внутри переменной части выражены переменные поля различных рубрик. Для каждой рубрики в определенных полях (два поля в начале элемента, начало каждой рубрики) сообщается имя рубрики и длина переменной информации. Библиографическая запись имеет среднюю длину 1500 символов.

Данная структура массива дает возможность быстрого доступа к библиографическим записям по рубрикам, минимального расхода памяти и обеспечения записей переменной длины.

В информационном отношении, как мы уже отметили, библиографические записи состоят из двух частей: фиксированная часть записи и переменная часть записи. Общий формат показан на рис. 3.

90	
фиксированная часть записи	переменная часть
I	90 91

Рис.3.

Фиксированная часть библиографической записи о документе (аннотация документа) состоит из 90 символов. Она имеет 15 логических полей, которые используются как характеристики для дополнительного поиска информации (поиск по характеристикам, не являющимся дескрипторами).

Рассмотрим кратко содержание этих логических полей (формат фиксированной части записи показан на рис. 4).

NFO	NOM	RN	TOR	BL	VV	IN	SC	LI	AU	CR	HR	FR	B	F
9	7	6	I	3	2	2	3	9	30	2	2	4	4	6

Рис.4.

NFO - 9 символов; номер поискового образа в ОМПОД, соответствующий данной библиографической записи (2 символа - номер диска, 3 символа - номер зоны на диске, 4 символа - номер элемента в зоне. Это первый элемент поискового образа документа, индекс которого является номером документа). Это поле служит для контроля и для создания обратной связи от МД к ОМПОД.

NOM - 7 символов; номер микрофиша (в библиотеке), соответствующий данному документу.

RN - 6 символов; это (для документов из фонда ИНИСа) reference number, являющийся номером библиографии в Atomindex.

TOR - 1 символ; показывает тип библиографической записи и имеет 9 значений: B, C, D, F, G, H, P, R, Y, (их значение см. в /2/).

BL - 3 символа; определяет библиографический уровень данной записи. Может быть комбинацией максимум из трех данных символов A, M, S, C, (см. тоже /2/).

VV - 2 символа; определяет том Atomindex, где записана библиография документа в ИНИС.

IN - 2 символа; определяет номер соответствующего издания.

SC - 3 символа; определяет предметную категорию документа (см. /2/).

L - 9 символов; литературный индикатор (см. /2/).

AU - 30 символов; автор публикации.

CR - 2 символа; сокращение страны.

YR - 2 символа; год издания.

IR - 4 символа; частота использования документа.

B - 4 символа; длина фиксированной + переменной частей записи.

F - 6 символов; поле для дополнительной информации.

В переменной части записи, являющейся строкой, последовательно входит вся оставшаяся информация, посылаемая ИНИС ом. Эта информация состоит из рубрик (100, 110, 111, 200, 201, 211, 213, 230, 231, 250, 300, 310, 401, 402, 403, 500, 600, 610, 700, 710).

Таким образом, при распечатке библиографии документа можно получить всю исчерпывающую информацию о нем.

Если не имеется в наличии достаточно дисков, этот массив может быть расположен, в данном виде на МД, но тогда, конечно, доступ до отдельной библиографии последователен. Данная структура принята как основная при создании архива информационного фонда на МД. В любой момент при реактивизации некоторого документа (при его перебросе из архива в оперативное хранилище) не требуется никакой трансформации структуры для осуществления его включения в массив с прямым доступом.

Как мы уже отметили, в массиве документов совершается дополнительный поиск по желанию потребителя по отдельным характеристикам (в фиксированной части библиографической записи), которые не являются дескрипторами. Данная организация массива МД имеет определенные преимущества при пакетной обработке запросов, поскольку ответы на всю группу запросов могут быть получены за один просмотр массива. Кроме того, если поиск производится по ключам различных типов (как в данном случае, например, по году публикации, фамилии автора и т.п.), последовательная организация массива является единственной практически реальной схемой, поскольку обычно не представляется возможным хранить многочисленные копии массива, удовлетворяющие различным

желаемым способом упорядочения. Некоторые математические соотношения времени поиска в таком массиве показаны в /3/.

#### Основной массив поисковых образов документов (ОМПОД)

Массив ОМПОД строится как отображение массива документов. Это означает, что индексы (адреса) первых членов узлов в ОМПОД должны находиться в соответствии с индексами (адресами) библиографии документов в массиве документов. Это соответствие может быть реализовано различными способами. В данном случае, например, оно осуществляется с помощью адресов связи, включенных в состав поисковых образов документов. Второй возможный способ заключается в том, что библиография каждого документа отводится в  $K$  раз больше ячеек, чем под его поисковый образ ( $K = \text{const}$ ). Причем, число  $K$  выбирается таким образом, чтобы объем памяти, отведенный под библиографию документов, был достаточным для этой цели.

ОМПОД строится следующим образом. Поисковым образом каждого документа является некоторый набор дескрипторов. Эти дескрипторы, сгруппированные вместе, образуют так называемый узел, характеризующий данный документ. Номер узла документа (т.е. индекс в основном массиве поисковых образов документов) в явном виде не указывается. Он совпадает с номером (индексом) того члена данного узла, в котором находится его первый дескриптор. Каждый узел документов имеет вид последовательности пятерок кодов, которую можно представить следующим образом:

I	$a_{i1}$	$b_{i1}$	$c_{i1}$	$d_{i1}$
0	$a_{i2}$	$b_{i2}$	$c_{i2}$	$d_{i2}$
"	"	"	"	"
0	$a_{ik-1}$	$b_{ik-1}$	$c_{ik-1}$	$d_{ik-1}$
2	$a_{ik}$	$b_{ik}$	$c_{ik}$	$d_{ik}$

Рассмотрим отдельные составные части каждого узла. Цифра в первой колонке представляет собой некоторый признак, принимающий одно из трех значений: 1, 0, 2. Значение 1 указывает начальную строку узла некоторого документа, 2 — последнюю строку, а значение 0 присутствует в каждой из промежуточных строк. Код  $a_i$  является кодом некоторого дескриптора, индексирующего данный документ;  $a_i$  есть номер элемента массива на магнитном диске, где расположена первая часть ОМД. Код  $b_i$  является адресом связи, т.е. это индекс того элемента в данной зоне, в котором в составе узла, ближайшего к данному, снова записан тот же дескриптор. Код  $c_i$  обозначает резервное поле для дополнительной информации.

Код  $d_i$  является адресом связи, устанавливающим соответствие между поисковым образом документа и его библиографией, находящейся в массиве документов. В данном случае это номер диска и номер элемента диска, с которого начинается библиография документа на МД.

Таким образом, для каждого поискового образа документа устанавливается единая форма записи переменной длины, а при помощи связи в цепном списке соединяются все поисковые образы данного дескриптора. В начало каждого дескрипторного списка попадаем с помощью соответствующего заголовка из массива МЗД (смотри дальше описание этого массива).

Следует заметить, что вся память, отведенная под ОМПОД, делится на зоны, в которых размещаются узлы, причем в каждой зоне должно быть размещено целое число узлов. Списки ОМПОД формируются в пределах одной зоны. Заметим, что один и тот же дескриптор может встречаться в списках разных зон. Другими словами, список для данного дескриптора сегментируется таким образом, что сегмент списка существует в каждой зоне, в которой встречается данный дескриптор, а

сегменты в явном виде в самом массиве не соединяются. Со структурно-функциональной точки зрения ОМПОД представляет следующую структуру (см. рис.5), расположенную на магнитном диске.

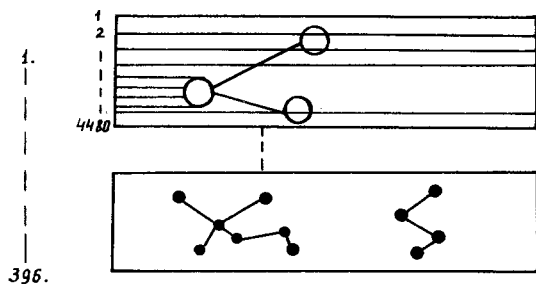


Рис.5.

Каждый элемент узла имеет следующий формат:

ПР	$a_i$	$b_i$	$c_i$	$d_i$
I	5	4	2	8

ПР - один символ ;

$a_i$  - 5 символов ;

$b_i$  - четыре символа ;

$c_i$  - два символа ;

$d_i$  - восемь символов.;

Каждый элемент, таким образом, занимает два слова в памяти на диске. Величина каждой зоны принята в 1/2 цилиндра дискового пакета (характеристика дискового пакета описана в /4/).

Выбор зоны величиной 1/2 цилиндра дает возможность использовать "естественную" сегментацию диска на цилиндры. Это имеет боль-

шое значение для процесса поиска, потому что для обработки данного цепного списка не требуется механического перемещения читающей головки.

Итак, на магнитном диске можно иметь 396 таких зон. Если в среднем на индексирование каждого документа уходит 10 дескрипторов, то в одной зоне можно иметь 448 поисковых образов, а на диске -  $448 \times 396 = 177408$  поисковых образов документов.

Методика программирования алгоритмов формирования структуры подобных массивов описана в /5/.

Этот массив заполняется позонно, в хронологическом порядке. Если в некоторой зоне ОМПОД в поисковом образе документов впервые встретился дескриптор, то для этого дескриптора "открывается" заголовок в МЗД.

Это означает, что элементами такого массива являются фиксаторы списков поисковых образов документов. Каждый заголовок (в дальнейшем будем употреблять слово "фиксатор") состоит из 6 кодов и имеет вид:

$$a \quad c_i \quad N_{ij} \quad d_j \quad e_j \quad c_j^i,$$

где  $a$  - поле для дополнительной информации (резервное поле);

$c_i$  - адрес связи, показывающий на следующий заголовок данного дескриптора, находящегося в той же зоне массива МЗД;

$N_{ij}$  - число документов, включенных в  $j$ -й список этого дескриптора;

$d_j, e_j$  - номера диска и зоны на диске, указывающие адрес списка;

$c_j^i$  - адрес первого документа списка.

Как мы уже отметили, в одной зоне массива МЗД могут встречаться заголовки различных списков для одного и того же дескриптора. С целью сокращения общего числа членов в массиве ОМД и улучшения метода поиска все заголовки данного дескриптора соединены в цепном спис-



ке. Сами эти списки заголовков сегментированы по зонам массива.

Таким образом, каждая строка МЗД является представителем некоторого дескриптора и связана с первой строкой, представляющей тот же дескриптор в зоне ОМПОД, отвечающей тому сегменту, который содержит упомянутую строку МЗД. Каждой зоне магнитного диска, содержащего ОМПОД, соответствует сегмент в МЗД.

Массив заголовков дескрипторов используется при поиске и представляет 2-ой уровень иерархии управляющей части системы поиска.

Со структурно-функциональной точки зрения МЗД представляет структуру, показанную на рис.6. Каждый элемент массива состоит из 20 символов и имеет следующий формат:

- a - четыре символа;
- c<sub>i</sub> - четыре символа;
- N<sub>ij</sub> - четыре символа;
- d<sub>j</sub> - один символ;
- e<sub>j</sub> - три символа;
- c<sub>j</sub> - четыре символа.

Последний элемент цепного списка отмечен признаком конца списка (КС).

Структурные связи элементов одной цепочки показаны на рис.6.

#### Основной массив дескрипторов ОМД

Этот массив является дескрипторным (ключевым) указателем. Он в свою очередь имеет списковую структуру и является первым уровнем иерархии управляющей части основного информационного массива (ОМПОД). В список объединены элементы, содержащие адреса отсылки к каждому первому элементу цепных списков заголовков одного и того же дескриптора (см. рис.6 и 7). Сами дескрипторы в ОМД в явном виде

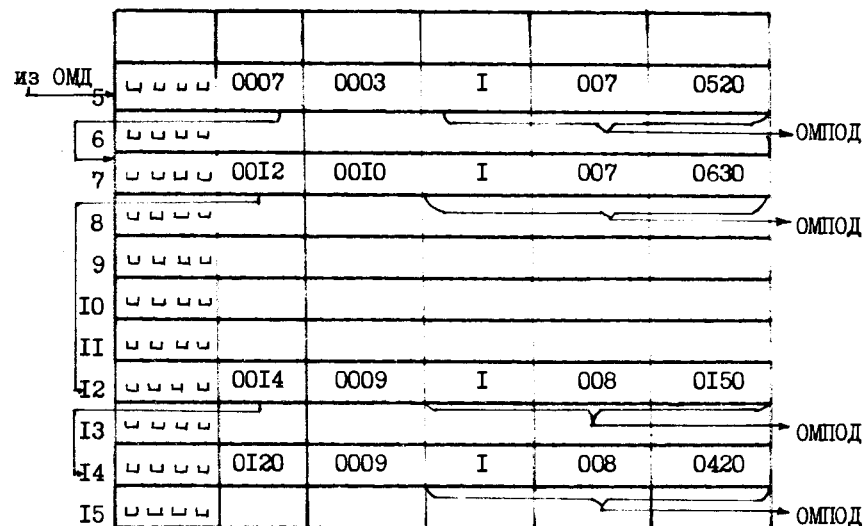


Рис. 6. Фрагмент зоны МЗД

не указываются. Они представлены в виде индексов (адресов) начал списков ОМД и их коды совпадают с индексами элементов первой части основного массива дескрипторов.

Остановимся на описании ОМД несколько подробнее. Он состоит из двух частей (см. рис.7). В первой части каждая ее строка соответствует определенному дескриптору и имеет связь:

- а) со строкой МЗД, которая является первой из соответствующих тому же дескриптору в данной зоне МЗД;
- б) со строкой, входящей во вторую часть ОМД, если упомянутому дескриптору в МЗД соответствуют строки, расположенные в разных зонах.

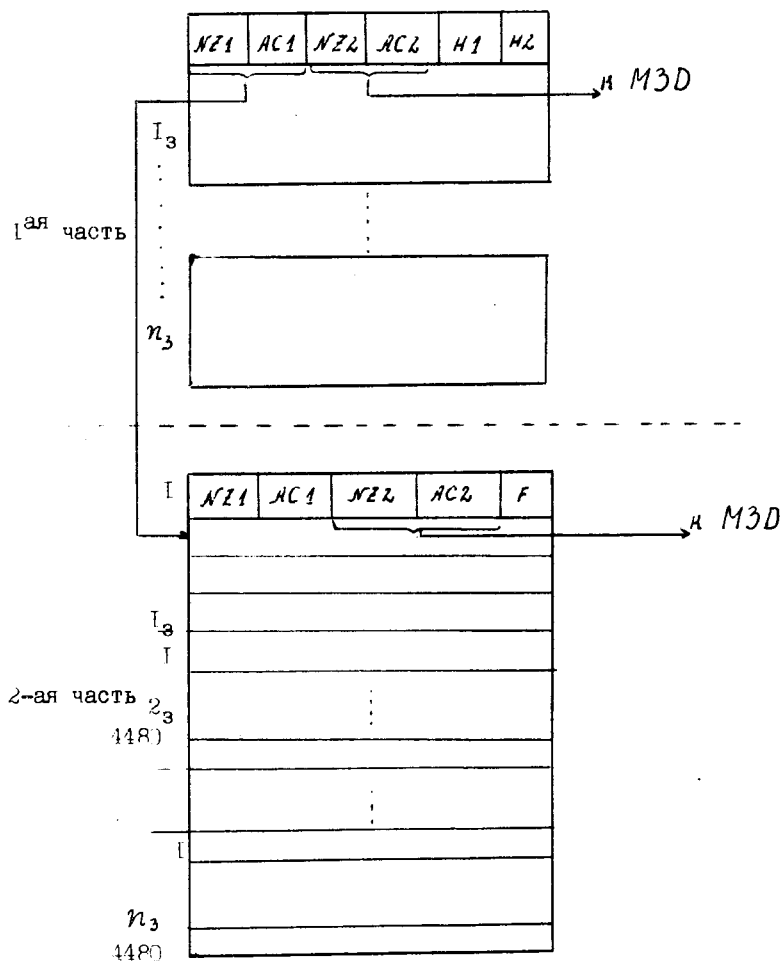


Рис.7.

Во второй части ОМД каждая строка соответствует одному из дескрипторов, который представлен либо строкой, содержащейся в первой части ОМД, либо строкой второй части ОМД, расположенной выше данной строки. Каждая строка второй части ОМД имеет связь:

а) со строкой некоторой зоны МЗД, отвечающей тому же дескриптору;

б) с расположенной ниже строкой второй части ОМД, если упомянутый дескриптор представлен строкой в некоторой зоне МЗД, с которой не связаны ни строки первой части ОМД, ни расположенная выше строка второй части ОМД.

Каждый элемент основного массива дескрипторов содержит следующую информацию (см. формат записи на рис.8).

NZ1	AC1	NZ2	AC2	H1	H2
3	7	3	4	3	3

Рис.8.

NZ1 , AC1 - это адрес следующего элемента списка ОМД, где NZ1 - это номер зоны, а AC1 - номер элемента внутри зоны;

NZ2 , AC2 - адрес заголовка в массиве МЗД;

H1 - число членов цепи определенного дескриптора в зоне NZ1;

H2 - число членов в цепи заголовков дескриптора в NZ2 .

Под клетками показано число символов, отведенных под каждое поле, в конкретной разрабатываемой ИПС. Последний элемент каждого списка рассматриваемого массива отмечен знаком конца списка (КС). Со структурно-функциональной точки зрения этот массив представляет структуру, показанную на рис.7. На рисунке изображены некоторые функциональные связи между элементами (и отдельными частями элементов), о которых уже было сказано.

Необходимо отметить, что весь массив ОМД логически разбит на отдельные зоны величиной в пол-цилиндра. Но цепные списки дескрипторов не сегментируются по зонам (поскольку их средняя длина намного короче, чем длина списков в ОМПОД). Здесь логическое деление на зоны связано с повышением эффективности стратегии поиска и с уменьшением числа перемещения головки диска при движении по спискам в режиме групповой обработки запросов.

Как видно, существенным элементом предложенной структурно-функциональной организации является понятие "зона". Подробнее об этом мы уже писали в /4/. Здесь остановимся конкретнее на сущности этого понятия.

Введение "зоны" решает в основном два конкретных вопроса:

- 1) зона является неотъемлемой частью процедуры поиска при определении степени смыслового соответствия стратегии поиска;
- 2) зона является "средством" совмещения процедуры доступа по спискам с управляемой длиной и дает возможность минимизации передвижения головки диска.

Правильное решение второго вопроса, со своей стороны, влияет на первый и увеличивает эффективность процедуры поиска. Так, напр., при оптимизации работы системы необходимо, чтобы данному дескриптору присваивалось минимальное количество зон. Это возможно, когда списки в каждой зоне формируются "плотнее", что достигается путем применения некоторой программы автоматической сортировки, которая реализует процесс оптимизации, направленный на размещение в зоне всех записей, относящихся к сильно связанным множествам дескрипторов.

В данном случае, например, при исследовании 10000 документов (они занимают 20 зон массива ОМПОД) можно подсчитать среднее коли-

чество документов в одной цепи. Используя формулу В.Верна /6/  
$$J_0 = \frac{F}{\ln F},$$
 где  $J_0$  - среднее количество документов на дескриптор,  $F$  - максимальная частота использования дескриптора, (в нашем случае  $F=1200$ ), получаем  $J_0 = 160$ . При наличии 20 зон, в каждой зоне в среднем можем иметь 8 членов в списке. Программа сортировки, конечно, не приводит к тому, что система будет отвечать на запросы с определенной комбинацией дескрипторов более эффективным образом, чем на запросы с любой другой комбинацией дескрипторов, а скорее ведет к тому, что любая комбинация дескрипторов в запросе становится равновероятной и оптимизирует поиск, основанный на произвольной взаимосвязи дескрипторов в запросе. Кроме того, чем выше плотность списков внутри зоны, тем меньше число зон, фигурирующих в списке дескриптора, что, со своей стороны, приводит к уменьшению длин массивов МЭД и ОМД.

Как мы уже отметили, величина зоны выбрана в размере половины цилиндра. Это означает, что обследование сегмента списка не требует перемещения считывающих головок, если они установлены у соответствующего цилиндра. Это имеет большое значение при групповой обработке запросов. Если бы каждый запрос обрабатывался отдельно, считывающая головка делала бы один цикл перемещения по цилиндрам, останавливаясь около тех из них, которые имеют отношение к исполняемому запросу. Однако, когда одновременно исполняется несколько запросов, считывающей головке необходимо сделать только один цикл перемещения по цилиндрам для целой группы запросов, хотя количество остановок при выполнении этого цикла перемещения увеличится.

При решении данного вопроса возникает интересная проблема, оказывающая большое влияние на определение величины массива, к которому необходим прямой доступ во время работы процедур поиска. Это, в свою очередь, решает вопрос о программной

реализации метода доступа и структуры массивов. Для ясности рассмотрим в качестве примера ОМПОД, в котором совершается последняя поисковая процедура при нахождении релевантных документов. Как уже отметили (см. рис.5), ОМПОД состоит из списков поисковых образов документов, сегментированных по зонам. При групповой обработке запросов необходимо обрабатывать параллельно, в общем случае, несколько списков в данной зоне. Суть поиска на этом этапе состоит в сравнении дескрипторов в узлах, формирующих данный список (см. рис.5). Эти списки расположены в определенной зоне на диске. Вопрос, который нам предстоит решить, выбирая данную зону (соответственно данный цилиндр) - как осуществить доступ к отдельным узлам; прочитать всю зону в ОП ЭВМ и потом обследовать нужные списки, или каждый раз записывать в память только определенный узел. Проиллюстрируем это соответствующими расчетами.

В /4/ приведены основные характеристики диска и операционной системы СДС при работе с дисками; описан метод заполнения отдельных цилиндров диска. Показано, что для заполнения половины цилиндра (140 секторов, 8960 машинных слов) необходимо сделать двадцать оборотов дискового пакета. Следовательно, чтобы осуществить прямой доступ к информации в одной зоне, необходимо время величиной  $t_s$  :

$$t_s = t_{\Pi} + t_o + t_{\text{тр}} \quad (1)$$

Как видно из формулы,  $t_s$  имеет три компоненты, где

$t_{\Pi}$  - это время, необходимое для нахождения данного цилиндра;

$t_o$  - время оборота дискового пакета;

$t_{\text{тр}}$  - время трансмиссии;

Для ЭВМ СДС-6200:

$$t_{\Pi} = 75 \text{ мс}$$

$$t_o = 25 \text{ мс}$$

Время трансмиссии (в миллисекундах)  $n$  символов определяется по формуле

$$t_{\text{тр}} = \frac{n}{\theta} \times 25 = \frac{n}{\theta} \times t_o,$$

где  $\theta$  - число символов на данной дорожке. В нашем случае  $t_{\text{тр}} = \frac{n}{5016} \text{ мс}$ ;

$n = 4508$ ;  $t_{\text{тр}} = 1/2 \text{ мс}$ . Когда читается вся зона в ОП,  $t_b$  - время доступа до информации в данной зоне определяется по формуле:

$$t_b = t_{\Pi} + 20t_o + 10t_o = t_{\Pi} + 30t_o$$

для нашего случая  $t_b = 75 + 500 + 250 = 825 \text{ мс}$ . Это означает, что при чтении всей зоны методом прямого доступа каждый раз будет расходоваться время  $t_b = t_{\Pi} + 30t_o$ . Но в этом случае не тратится время на задержку при чтении данной дорожки, т.е. всегда читается информация всей дорожки. Для выяснения этого вопроса рассмотрим проблему чтения информации на дорожке цилиндра. Пусть информация занимает половину дорожки (см. рис.9).

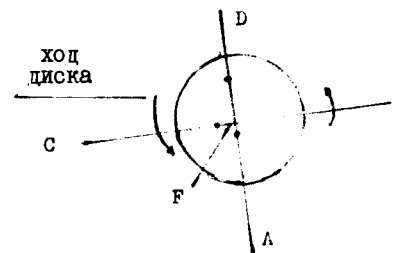


Рис.9.

Если чтение начинается тогда, когда головка находится в точке А, не будет никакой задержки, и общее время для чтения в этом случае ограничивается временем трансмиссии; для данного примера это  $(\frac{1}{2} \times 25) \text{ мс}$ .

Если чтение начинается при нахождении в точке В, имеется 1/4 задержки и 1/2 для трансмиссии, в общем,  $(3/4 \times 25) \text{ мс}$ .

Если читающая головка находится в точке Д, имеется 1/2 задержки и 1/2 трансмиссии - в общем - 25 мс.

Интересный случай - головка находится в С. В будущей ЭВМ, вероятно, можно будет чтение начать сразу, во второй части входного буфера. Потом, после задержки (1/2 x 25) мс, запись будет совершаться в первой части буфера и заканчиваться, когда т.С попадает снова под магнитную головку. Таким образом, общее время (задержка + трансмиссия) никогда не превзойдет времени одного оборота, независимо от начальной позиции считывающей головки. Когда читается целая дорожка, не имеется никакой задержки. Максимальная задержка при чтении - это время одного оборота, если головка находится в точке F<sub>0</sub>, расположенной наиболее близко к точке А, но не совпадающей с ней.

В общем случае, когда необходимо совершить прямой доступ к некоторому количеству узлов в данной зоне (после определения цилиндра), мы не знаем, в какой позиции находится считывающая головка. Так как, вообще говоря, эти узлы могут быть разбросаны совершенно произвольно в зоне, то это означает, что они находятся в случайных местах на дорожках данного цилиндра. Тогда не всегда оказывается эффективной запись каждого узла отдельно в ОП, т.е. из-за времени задержки может случиться, что при доступе к узлам расходуется времени больше, чем на 20 оборотов дискового пакета.

Следовательно, структурно-функциональная организация массивов должна быть такой, чтобы на основе количества обрабатываемых узлов (или элементов для массива МЗД и ОМД), можно было определять ту часть зоны массива, которую нужно прямым доступом читать в ОП ЭВМ.

При программной реализации необходимо так формировать соответ-

ствующие массивы, чтобы отдельные элементы, к которым можно осуществлять прямой доступ в разных массивах, были (соответственно) следующими:

- 1) В массиве документов - одно машинное слово.
- 2) В ОМПОД - элемент узла.
- 3) В ОМД и МЗД - заголовки дескрипторов.

Для программной реализации подобной структуры можно выбрать относительную (Relative) организацию массива прямого доступа. При этом экономится значительное место в памяти, т.е. ключ для прямого доступа в явном виде не указывается; кроме того, массив можно читать как выборочно (с прямым доступом), так и последовательно, что дает возможность менять величину выбранной для чтения информации.

Структурно-функциональная организация основных поисковых массивов является достаточно гибкой и адаптивной как в функциональном, так и в структурном отношении, что дает возможность оптимально применить выбранную стратегию поиска релевантных документов в большом информационном массиве.

## ЛИТЕРАТУРА

1. Д.Д.Арнаулов. "Об одном способе организации информационного массива и дескрипторного словаря в библиографической ИПС". Сб. Цифровая вычислительная техника и программирование, вып.7, 1970, "Советское радио", М.
2. Д.Д.Арнаулов, В.А.Бирюков. "Основные характеристики информационной системы ИНИС". Депонированная публикация ОИЯИ, Б4-11-8553, Дубна, 1975.
3. Д.Д.Арнаулов. "Анализ методов доступа информации к внешнему ЗУ на магнитных дисках при работе ИПС, реализованной на ЭВМ третьего поколения. Препринт ОИЯИ, IO-7953, Дубна, 1974.
4. Д.Д.Арнаулов. "Выбор оптимальной структуры основного информационного массива ИПС для размещения на магнитном диске". Сообщение ОИЯИ, IO-7949, Дубна, 1974.
5. Д.Д.Арнаулов. "Организация мультисписочных узловых структур и их программная реализация на КОБОЛе. Сообщение ОИЯИ, IO-7587, Дубна, 1973.
6. В.Верн . "Характеристика ИПС", сб. "Информационный поиск", М., 1970, "Советское радио".

Рукопись поступила в издательский отдел  
19 февраля 1975 г.