

СООБЩЕНИЯ  
ОБЪЕДИНЕННОГО  
ИНСТИТУТА  
ЯДЕРНЫХ  
ИССЛЕДОВАНИЙ

ДУБНА



6/II-78

710/2-78

10 - 11051

Н.Н.Говорун, Ю.П.Залаторюс, С.М.Мальцева,  
В.И.Никитина, В.М.Сумароков, Г.Н.Тентюкова

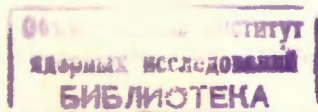
ОРГАНИЗАЦИЯ ИНФОРМАЦИОННЫХ МАССИВОВ  
СИСТЕМЫ "КАДРЫ"

1977

10 - 11051

Н.Н.Говорун, Ю.П.Залаторюс, С.М.Мальцева,  
В.И.Никитина, В.М.Сумароков, Г.Н.Тентюкова

ОРГАНИЗАЦИЯ ИНФОРМАЦИОННЫХ МАССИВОВ  
СИСТЕМЫ "КАДРЫ"



Говорун Н.Н. и др.

10 - 11051

Организация информационных массивов системы "КАДРЫ"

Рассматриваются вопросы организации информационных массивов системы "КАДРЫ": разнообразие признаков интерпретаций машинных образов объектов, структуры признаков записей, динамическое распределение машинной памяти и др.

Система реализована на ЭВМ БЭСМ-6.

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1977

Govorun N.N. et al.

10 - 11051

Organisation of the "Personnel" System Information Files

Organisation of the "Personnel" System Information Files is considered: variety of feature interpretation of machine images of objects, the structures of feature articles, the dynamic distribution of machine memory etc. The system is realised with the BESM-6 computer.

The investigation has been performed at the Laboratory of Automation, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna 1977

Автоматизированная информационная система "КАДРЫ" предназначена для накопления, хранения, обновления, обработки, поиска и выдачи сведений по кадровому учету в ОИЯИ.

Одной из важнейших проблем при разработке любой автоматизированной информационной системы является проблема организации информационных массивов (файлов), от решения которой зависит эффективность использования ЭВМ. Трудности решения задачи организации файлов обусловлены противоречием между большим объемом информации и ограниченными ресурсами ЭВМ. Организация информационных массивов системы включает выбор формата для записи элементов как массива сведений, так и служебных массивов, закрепление и распределение типов файлов в устройствах внешней памяти ЭВМ, способ разбиения файлов на подфайлы для эффективной работы поисковой системы. Структура хранения в сильной степени зависит от структуры данных.

Структура данных

Основным структурным компонентом массива сведений об объектах системы является "анкета", где указывается имя объекта и значения его признаков.

Признаки, характеризующие объект системы, могут быть отнесены к одному из трех типов: простой, сложный и сложный списковый.

Простой признак выражается элементом \*) фиксированной длины. Сложный несписковый признак выражается простой группой \*\*), состоящей из элементов фиксированной или переменной длины. Сложный списковый признак выражается совокупностью повторяющихся простых групп. Элементы, составляющие сложный признак, называются подпризнаками. "Анкета"\*\*\* - главная структурная единица для представления объекта.

Простые признаки задают обязательные свойства объекта и размещаются в фиксированных полях (например, пол, табельный номер). Сложные признаки соответствуют свойствам, которыми обладают не обязательно все объекты системы (например, данные о научном сотруднике), или свойствам, обязательным для объекта, но требующим размещения в переменном поле (например, домашний адрес). Сложные списковые признаки - это сложные признаки, которые могут иметь множественные значения (например, данные о членах семьи). Особым признаком анкеты является дата последнего изменения. Она заносится автоматически при вводе или редактировании анкеты.

Значение простого признака или подпризнака может быть выражено числом, строкой символов или датой. Признак или подпризнак обладает индикатором незадействованности: если его значение неопределено, то ставится специальная метка. Сложный признак может иметь только одно из трех типов значений:

0 означает, что объект не обладает признаком,

— означает отсутствие данных о значении признака.

Любое другое значение указывает на наличие содержательного значения признака, которое задано значениями соответствующих подпризнаков.

Каждый признак или подпризнак имеет номер и мнемоническое имя, задаваемое строкой символов. Признаки (подпризнаки) принадлежат к одному из двух типов представления: кодированный и описательный. Кодированный - признак, который допускает конечный, заранее определенный набор значений, каждому из которых сопоставлен код (шифр). Описательный - признак, для которого нельзя или нецелесообразно заранее определить полный набор значений.

\*) Элемент - элементарная, логически не делимая единица данных, из которой в конечном счете составляются все остальные типы структур данных. Простая группа - совокупность элементов /I/.

\*\*\*) Далее в тексте термин "анкета" употребляется без кавычек.

Кодированный признак всегда имеет числовое значение. Значение описательного - строка символов, дата или число.

Для каждого признака в системе заданы: длина соответствующего поля (количество символов или разрядов) и формат представления (число, дата или строка символов). Для кодированного признака также задается список его дискретных значений и полные текстовые интерпретации значений.

К специальным свойствам признаков относятся:

- возможность использования со служебными словами "год", "возраст", "стаж";

- блокировка доступа (замок защиты);

- блокировка активности (временное изъятие признака из списка).

Свойства признаков зафиксированы в специальных системных таблицах и используются для контроля сведений, поступающих в базу данных, для формирования базы данных, для контроля правильности запроса, для распаковки значений признаков и распечатки сведений.

#### Структура хранения

Пользователь готовит данные в виде строк символов. После ввода в ЭВМ данные преобразуются и заносятся во внешнюю память для хранения.

Структура хранения обуславливается тем, как будут использоваться эти данные, а также характеристиками запоминающих устройств ЭВМ и возможностями операционной системы. На организацию структуры хранения оказывает влияние также структура данных.

Каждая анкета заносится в файл анкет в порядке поступления, первоначально в виде отдельной физической записи. Данные, относящиеся к каждому сложному списковому признаку, образуют гнездо<sup>/2/</sup>.

Запись анкеты состоит из двух частей:

- 1) фиксированного поля, содержащего значения простых признаков и адреса значений сложных признаков;
- 2) переменного поля, содержащего значения сложных признаков, элементы которых записаны с постоянным или переменным форматом.

Значения признаков запоминаются в поле анкеты в соответствии со специальными таблицами, устанавливающими отношение между именем признака и положением его значения в поле записи.

В качестве имени объекта выступают фамилия, имя и отчество (ФИО) сотрудника, а также уникальный номер анкеты (АН), присвоенный ей при первоначальном вводе. Первое имя (ФИО) обусловлено характером объектов системы "КАДРЫ". Поскольку система имеет дело с анкетами сотрудников учреждения, то естественно в качестве имени объекта иметь ФИО сотрудника. Но идентифицировать объект только по этому имени не всегда удобно, во-первых, из-за его громоздкости, а во-вторых, из-за отсутствия гарантии уникальности. Поэтому ссылка на имя объекта внутри системы (а в некоторых случаях и извне системы) производится по второму имени- номеру анкеты.

В целях ускорения поиска анкеты по имени ФИО это имя запоминается системой в усеченном и преобразованном виде. Под имя отводится 11 символов фамилии и инициалы, которые записываются в специальном шеститбитном коде.

Таким образом, информация об объектах системы после ввода в ЭВМ расчленяется на основную (значения признаков) и имя. Основная заносится в файл анкет, а оба имени (ФИО и АН) - в таблицу ФИОАН. Одновременно составляется список адресов анкет (АНАДР) и таблица обратного соответствия фамилии номеру объекта (АНФИО). Затем все эти данные заносятся во внешнюю память ЭВМ для последующего использования при поиске и выдаче информации из системы.

В процессе работы система допускает дополнение данных и, следовательно, расширение области записи. Это расширение возможно за счет заполнения резерва памяти, выделенного для каждой анкеты (в этом случае анкета сохраняется в виде отдельной физической записи), или, при исчерпании резерва, анкета целиком переносится на свободное место файла, образуя "дыру", или же происходит формирование новых гнезд. В последнем случае запись одной анкеты оказывается составленной из нескольких фрагментов и, следовательно, время выборки анкет из внешней памяти может заметно возрасти. В этих случаях производится процесс упорядочения памяти, который сводится к сборке анкет по фрагментам, размещению их в порядке возрастания номеров анкет и ликвидации "дыр".

Необходимость выполнения процесса упорядочения памяти возникает тем реже, чем больше размеры резервных областей. С другой стороны, расширение размеров резерва нежелательно из-за ограниченного объема внешней памяти и некоторого увеличения времени

выборки информации из внешней памяти при удлинении анкет за счет резерва. Для разрешения этого противоречия требуется уделить особое внимание в процессе эксплуатации системы подбору констант, устанавливающих размеры резервных областей.

#### Способ организации основного файла

Организация файлов должна обеспечить быстрый доступ к записям файлов как при выдаче справок, так и при обновлении базы данных.

Поскольку система работает в оперативном режиме, что требует доступа в реальном масштабе времени, то очень важным является вопрос времени ответа на запрос. Эта характеристика системы в большой степени зависит от способа организации файлов, главным образом от способа разбиения файла на логические подфайлы.

Структура основного файла, содержащего анкеты, допускает поиск внутри файла по составному логическому ключу /3/. В качестве ключей система использует возможные значения поисковых признаков, т.е. признаков, по которым целесообразно вести поиск. Ключи, являющиеся групповыми, соответствуют не одному значению признака, а интервалу из допустимого множества значений. Все ключи внесены в Справочник ключей.

В системе "КАДРЫ" применен способ организации основного файла методом инвертированного списка /3/. Для каждого ключа формируется множество номеров анкет, содержащих значение признака, соответствующее ключу. Это множество отображается в некоторую область памяти, рассматриваемую как непрерывная память, где каждый двоичный разряд соответствует анкете в порядке возрастания номеров анкет. Полученные таким образом "линейки" разбирают основной файл на подфайлы, давая возможность быстрого поиска множества анкет, релевантных запросу.

Строка Справочника ключей содержит mnemonic имя ключа (не больше 16 символов), его числовое представление, адрес "линейки", количество анкет, содержащих данное значение признака, и некоторую другую информацию. Массив Справочника ключей упорядочен по именам ключей относительно букв русского алфавита.

## Организация внешней памяти

В качестве внешней памяти используются запоминающие устройства с прямым доступом - дисковые пакеты ЕС-5052, а способ доступа определяется возможностями операционной системы "Дубна" /4,5/. Все массивы данных: основной файл, Справочник ключей, системные таблицы, тексты, "линейки" и другие располагаются на одном дисковом пакете \*).

Весь пакет разделен на четыре участка /4/, длина каждого не превышает 256 трактов (зон). Любой вид информации (анкеты, "линейки", таблицы и т. д.) формируется и записывается во внешнюю память, а затем считывается постранично (1024 слова). Внешний адрес записи любого вида информации задается двадцатиразрядным числом: номер участка (2 разряда), номер тракта (зоны) в участке (8 разрядов), место на тракте (10 разрядов). При записи информации на диск в его каталог заносится значение идентификатора вида информации - число от 1 до 7.

Таблицы, которые образуются в процессе формирования основного файла (ФИОАН, АНАДР, АНФИО), имеют значительные объемы и наращиваются по мере расширения фонда анкет. В целях оптимального использования внешней и основной памяти запись на МД таких таблиц производится страницами в любую (очередную) свободную зону диска, определяемую по каталогу, так что физически каждая такая таблица может оказаться разделенной на участки, занимающие несколько трактов (зон). Вызов таблиц с диска в основную память выполняется в разных режимах в зависимости от потребности программы и наличия памяти:

- вызов всей таблицы,
- вызов зоны по номеру,
- вызов зоны, содержащей запись с заданным ключом,
- вызов зоны, содержащей запись с заданным порядковым номером.

Такую же организацию на диске имеют те системные таблицы, которые требуют больших объемов памяти. Текущее состояние на внешней памяти таких таблиц отражено в массиве статусной информации ХАДТ .

\* В случае нехватки памяти на диске предусмотрена возможность записи части данных на магнитную ленту.

## ЛИТЕРАТУРА

1. Информационные системы общего назначения (аналитический обзор систем управления базами данных). Москва, "Статистика", 1975.
2. А.И.Китов. Программирование экономических и управленческих задач. Москва, "Советское радио", 1971.
3. Д.Лефковиц. Структуры информационных массивов оперативных систем. Москва, "Советское радио", 1973.
4. В.Ю.Веретеннов и др. Дисковая операционная система. ИАЭ-2486, Москва, 1975.
5. Г.Л.Мазный. Вариант подпрограмм для обмена с МД, МБ и МД строками и страницами на ФОРТРАНе. Сообщение ОИИИ, ИИ-9845, 1976.

Рукопись поступила в издательский отдел  
1 ноября 1977 года.