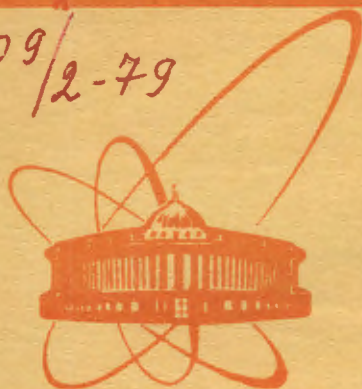


4509/2-79



сообщения
объединенного
института
ядерных
исследований
дубна

Ц840Г

Г-577

5/II-79

P10 - 12446

Н.Н.Говорун, С.М.Мальцева, В.И.Никитина,
Г.Н.Тентюкова

ФОРМИРОВАНИЕ, ОБНОВЛЕНИЕ

И РЕДАКТИРОВАНИЕ БАЗЫ ДАННЫХ ИСК

1979

P10 - 12446

Н.Н.Говорун, С.М.Мальцева, В.И.Никитина,
Г.Н.Тентюкова

ФОРМИРОВАНИЕ, ОБНОВЛЕНИЕ
И РЕДАКТИРОВАНИЕ БАЗЫ ДАННЫХ ИСК

ОБРАЗОВАТЕЛЬНЫЙ ИНСТИТУТ
НАСРНЫХ ИССЛЕДОВАНИЙ
БИБЛИОТЕКА

Говоруи Н.Н. и др.

P10 - 12446

Формирование, обновление и редактирование базы данных ИСК

Рассмотрены вопросы формирования, обновления и редактирования данных в системе управления базой данных ИСК, предназначенной для работы с документами картотечного типа и реализованной на ЭВМ БЭСМ-6. Дано описание входных форматов и системы синтаксического и смыслового контроля вводимой информации, рассмотрены управляющие карты для формирования пакетов заданий на ввод и редактирование документов. Изложен способ формирования основного информационного массива. Указаны типы признаков, характеризующих документ, содержащий набор сведений об объекте системы. Система ориентирована на использование в информационно-поисковых системах и реализована в ОИЯИ для решения задач кадрового учета. Она может быть использована в других задачах АСУ, имеющих документы картотечного типа.

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Сообщение Объединенного института ядерных исследований, Дубна 1979

Govorun N.N. et al.

P10 - 12446

The Loading Actualization and Updating of Data Base ISK

The questions of loading, actualization and updating of data are considered for data base management system ISK, which is intended for card type documents handling and realized with BESM-6 computer. Data entry formats, syntactic and semantic rules for input information are given; control cards to form the job decks for loading or updating documents are considered. The way to form the main information file is given. Feature types are pointed out, which characterize the document, containing the collection of facts about the system object. The system is oriented to use in the information retrieval systems. It was used in JINR to solve the personnel problems. It may be used in other managements problems with card type documents.

The investigation has been performed at the Laboratory of Computing Techniques and Automation.

Communication of the Joint Institute for Nuclear Research. Dubna 1979

Система управления базой данных ИСК*) предназначена для работы с документами картотечного типа и реализована на БЭСМ-6.

Формирование, обновление и редактирование базы данных ИСК выполняется системой программ SYSDATA, входящей в систему управления базой данных.

Основной структурной единицей базы является документ, содержащий набор сведений об объекте системы.

Документ может иметь до 8192 признаков. Часть из них - сведения о текущем состоянии объекта, часть - история (архив). За каждым признаком закреплен определенный номер, называемый ключевым словом, или ключом.

Признаки, характеризующие документ, могут быть отнесены к одному из трех типов:

- 1) простой признак,
- 2) сложный признак,
- 3) сложный списковый признак.

Простым признаком называется признак, занимающий фиксированное поле и не делящийся на составные части.

Сложным признаком называется групповой признак. Он состоит из одного или нескольких подпризнаков, занимающих фиксированное поле или поле переменной длины.

Сложный списковый признак может состоять из нескольких записей, имеющих одинаковую структуру^{1/}.

В целях простоты, удобства редактирования и более тщательной проверки документов, вводимых в базу данных, формирование и редактирование базы выполняется в два этапа с использованием двух дисковых пакетов типа ЕС-5052, а способ доступа определяется возможностями доступа операционной системы "Дубна"^{3/}.

На вспомогательном диске накапливаются документы и фрагменты редакций документов в том виде, в каком они вводятся в ЭВМ.

*) ИСК - информационно-поисковая система для документов картотечного типа.

На базовом диске формируются основной информационный массив документов в компактном виде и вспомогательные таблицы для поисковой системы^{1,2,4,5/}.

Система допускает редактирование документов, содержащихся на базовом диске, а также фрагментов редакций и документов, накопленных на вспомогательном диске.

I. Накопление данных на вспомогательном диске

Система настраивается на определенный тип логического файла, указанный на управляющей карте.

Документы или фрагменты редакций документов, введенные с внешних носителей, проверяются программами контроля и записываются текстовыми строками^{3/} в один физический (временный) файл на вспомогательном диске в форме, соответствующей фиксированному формату.

Одновременно с этим система создает для заданного логического файла несколько списков, каждый из которых соответствует одному типу обрабатываемой информации, указанному на управляющей карте.

В настоящее время допустимы четыре списка для каждого логического файла: "исправить", "заменить", "убрать", "новые". Имена этих списков говорят о типе информации, хранящейся в них.

Структура всех списков одинакова. Каждая запись занимает два машинных слова. В первом слове хранится номер документа, во втором — адрес документа или фрагмента редакции документа во временном файле.

В списках "исправить", "убрать", "заменить" накапливаются номера документов, присвоенные им при вводе в базу данных.

Номера должны быть указаны на управляющих картах, предшествующих вводимым документам или фрагментам редакций.

Список "новые" содержит порядковые номера документов, присваиваемые им в момент засылки документа во временный файл.

В дальнейшем они заменяются свободными номерами из базы данных.

Форматы входных данных

Входные данные могут быть введены с использованием языка фиксированных позиций (фиксированный формат), языка ключевых слов (произвольный формат) и смешанного языка фиксированных позиций и ключевых слов (смешанный формат)^{2/}.

Смешанный формат характеризуется тем, что в нем часть сложных признаков (или все) могут быть введены в фиксированном формате, в то время как простые признаки — только в произвольном.

Формат ввода указан на управляющей карте, задаваемой для каждого документа или группы документов.

Фиксированный формат характеризуется жестким форматом ввода: каждый признак документа занимает фиксированную позицию входного носителя. Формат ввода задается системными таблицами.

Надо сказать, что фиксированные позиции занимают только простые признаки и подпризнаки внутри сложных признаков. Порядок задания самих сложных признаков произволен.

Произвольный формат имеет форму последовательности пар "признак-значение". Признак задается его ключом, а значение — строкой символов.

Пара заканчивается разделителем, а весь фрагмент ввода — признаком конца.

Порядок задания пар и их положение на входном носителе внутри фрагмента ввода произвольны.

В качестве разделителей в произвольном формате употребляются символы: "=", ",", ":", ".", "()".

Форматы пар "значение-признак"

1. Для простого признака:

$$N_{\text{пр}} = \text{пппп},$$

2. Для сложного признака:

$$N_{\text{пр}}. N_{\text{подпр}} = \text{пппп},$$

3. Для сложного спискового признака:

$$N_{\text{пр}}. N_{\text{подпр}} (N \text{ зап}) = \text{пппп},$$

Принятые обозначения в этих форматах:

N — номер,

{ пр — признак,

{ подпр — подпризнак,

{ зап - запись,
пппп - строка символов, обозначающая значение признака.

Разделитель ":" означает признак конца информации на входном носителе, если используется неполная строка носителя.

Некоторое ограничение на применение форматов ввода накладыва-ется при частичном редактировании документов. В этом случае при-меняется только произвольный формат, но с некоторым расширением:

1. Добавление новой записи к списку:

№пр. №подр (0) = пппп,

2. Удаление записи с номером N из списка:

№пр (N зап) = \$,

3. Удаление списка:

№пр (0) = \$,

В трансляторе с языка ключевых слов для хранения вводимой инфор-мации используются списки с двумя связями. Результат работы транс-лятора - получение документа в фиксированном формате. Для эконо-мии времени упаковка значений признаков совмещена с контролем.

Контроль

Контроль вводимой информации осуществляется за несколько про-смотров с использованием системных таблиц.

О каждой ошибке и предупреждении выдается диагностика. Пре-дупреждения не влияют на процесс дальнейшей обработки документа. При обнаружении ошибки документ проверяется до конца для выявле-ния последующих ошибок, выдается на печать вместе с diagnosti-ческими посланиями и в базу данных не вводится.

Первый просмотр

Проверяется синтаксис языка ключевых слов, если информация вводится в произвольном или смешанном форматах. Создаются табли-цы для транслятора и списки значений признаков.

Для всех форматов ввода создаются вспомогательные таблицы, проверяется правильность написания управляющих карт и правильность составления фрагмента ввода.

При вводе в фиксированном формате проверяется количество пер-фокарт, занятых фиксированной частью документа и каждой записью сложного признака.

При обнаружении ошибок на этом этапе документ дальше не обрабатывается, а выдается на печать вместе с диагностикой.

Второй просмотр

Синтаксический контроль значений признаков. Он включает:

1. Проверку соответствия количества символов в значении признака с заданным в таблицах.
2. Проверку правильности написания символов.

При обнаружении ошибок на этом этапе документ подвергается смыс-ловому контролю. Обнаруженные ошибки накапливаются в буфере оши-бок.

Третий просмотр

Смысловый контроль значений признаков.

Смысловый контроль выполняется полностью, если документ удачно прошел синтаксический контроль. В случае, если были за-фиксированы ошибки, производится частичный контроль в соответствии с указаниями системных таблиц.

Частичный контроль применяется также при частичном редактиро-вании документа. Во всех случаях значения признаков проверяются по допустимому интервалу значений или по дискретному набору допусти-мых значений. Особому контролю подвергаются даты и значения при-знаков, выступающих в роли связей между логическими файлами.

При редактировании документа дополнительно проводится контроль правильности имени документа. При выдаче фрагмента редакции до-кумента на печать дополнительно выдается его имя из базы данных.

Массив входных данных

Массив входных данных состоит из фрагментов ввода, каждый из которых заканчивается управляющей картой "конец", а весь мас-сив заканчивается управляющей картой "FINISH".

Фрагмент ввода - это один документ или фрагмент редакции одного документа. Каждому фрагменту должны предшествовать управ-ляющие карты (одна или несколько), содержащие имя логического файла, для которого вводится или редактируется документ, тип вход-ного языка, тип редакции и признак, указывающий, на каком диске, основном или вспомогательном, выполняется редактирование.

Управляющие карты, предшествующие фрагменту ввода, называ-ются внутренними управляющими картами. Их влияние локализовано

внутри фрагмента ввода. Существуют два типа внутренних управляющих карт:

- 1) для ввода нового документа;
- 2) для редактирования ранее введенных.

Имеют место еще так называемые особые управляющие карты. Они выполняют роль разделителей (например, "FINISH", "конец"). Формат управляющих карт в массиве входных данных произволен, а имена карт допускают сокращения до количества символов, заданных при генерации варианта системы. Массив входных данных может состоять только из карты "FINISH".

Формирование и ведение информационного массива на базовом диске

Набор внешних управляющих карт

Карты, управляющие всем процессом создания, ведения и редактирования базы данных, называются внешними управляющими картами. Они объединены в набор внешних управляющих карт, который заканчивается картой "конец".

Работа SYSDATA идет до тех пор, пока не исчерпаются все управляющие карты в наборе внешних управляющих карт.

Предусмотрена возможность искусственного прерывания работы с пульта ЭВМ. В наборе может быть несколько карт с одинаковым именем. Имена карт допускают сокращения.

Параметры управляющих карт задаются в фиксированных полях, но набор указателей внутри поля и их расположение произвольны.

Внешние управляющие карты настраивают систему на указанное имя логического файла и задают соответствующий блок обработки. По назначению внешние управляющие карты разделяются на сервисные, подготовительные и непосредственно управляющие процессом создания и изменения базы данных.

В целом вся задача по ведению базы данных делится на ряд независимых задач:

1. Подготовка вспомогательного диска к работе, ввод новых документов и фрагментов редакций, проверка их по тестам и запись во временный файл на вспомогательном диске.
2. Выдача на АЦПУ информации из временного файла.

3. Переброска новых документов из временного файла в базу данных в сжатом виде и создание соответствующих таблиц.
4. Редактирование и изменение документов в базе данных с использованием фрагментов редакций на вспомогательном диске.
5. Выдача на АЦПУ документов из базы данных.

Допускается работа каждой задачи независимо и в любом наборе, но последовательность их исполнения будет в описанном выше порядке.

Упаковка документов в базу данных, редактирование и изменение документов в базе данных

Документы или фрагменты редакций документов из временного файла по одному подаются в базу данных для упаковки или редактирования.

Упаковка производится с помощью системных таблиц. Одновременно создаются таблицы для поисковой системы для каждого логического файла, задействованного в ИСК.

Каждый документ заносится в базу данных первоначально в виде отдельной физической записи. Данные, относящиеся к каждому сложному списковому признаку, образуют гнездо.

Запись документа состоит из двух частей:

- 1) фиксированного поля, содержащего значение 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 документа в соответствии со специальными таблицами, устанавливающими соотношения между именем признака и положением его значения в поле записи.

В записи документа оставляется резерв на случай расширения документа.

Информация об объектах системы (документах) после ввода в базу данных расчленяется на основную (значения признаков) и имя. Основная заносится в основной информационный массив, а имя - в специальные таблицы. Для гарантии уникальности имени документа и для удобства использования (так как имя может быть громоздким) документу присваивается второе имя - номер.

В большинстве случаев ссылка на имя объекта производится по второму имени - номеру. Поэтому составляется таблица обратного соответствия имени номеру объекта.

Одновременно создается список адресов документов и таблицы связей между логическими файлами, задействованными в системе.

При обнаружении ошибок и некорректностей во время упаковки документа выдается диагностика и документ в базу данных не вводится.

Система проверяет, не является ли документ дублем документа, уже имеющегося в базе данных или во временном файле. При обнаружении "двойника" выдается диагностика и документ в базу данных не вводится.

При добавлении новых записей к сложному списковому признаку создаются новые гнезда в поле документа и документ записывается на старое место. Если резерв исчерпан, то документ целиком переносится на свободное место файла документов.

При обычном редактировании документ записывается на старое место. В случае необходимости выполняется редактирование связей между файлами.

Авторы выражают глубокую благодарность Н.С.Новиковой за помощь в создании программного обеспечения системы, а также Л.Э.Мальцевой и А.Л.Приходько за помощь в отладке программ.

Литература

1. Н.Н.Говорун и др. Сообщение ОИЯИ, 10-11051, Дубна, 1977.
2. Н.Н.Говорун и др. Сообщение ОИЯИ, 10-11052, Дубна, 1976.
3. Г.Л.Мазный. ОИЯИ, 11-9845, Дубна, 1976.
4. Н.Н.Говорун и др. ОИЯИ, 10-10951, Дубна, 1977.
5. Н.Н.Говорун и др. ОИЯИ, 11-10-10800, Дубна, 1977.
6. Д.Кнут. Искусство программирования для ЭВМ, т.1. Основные алгоритмы. Перевод с английского, изд-во "Мир", Москва, 1976.

Рукопись поступила в издательский отдел
7 мая 1979 года.