

3-224

На правах рукописи

ЗАЛАТОРЮС ЮОЗАС ПЯТРОВИЧ

**ВОПРОСЫ МАТЕМАТИЧЕСКОГО
ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННО-
ПОИСКОВЫХ СИСТЕМ**

Специальность: 01.01.10 - математическое обеспечение
вычислительных машин и систем

А в т о р е ф е р а т
диссертации на соискание ученой степени
кандидата физико-математических наук

ДУБНА - 1982

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы. Прогресс современной науки все в большей степени зависит от эффективности использования информации. Увеличиваются объемы научных исследований и лавинно возрастает поток научных данных. Кроме того, непрерывно усложняются структура и задачи систем управления. Появляются разного рода информационные задачи, которые невозможно решить традиционными способами. Это явилось причиной глубокого проникновения вычислительной техники во все сферы человеческой деятельности, связанные с производством и управлением. О степени развития государства уже судят не только по его промышленному потенциалу (и даже не столько), но и по способности обрабатывать информацию.

Использование для хранения, поиска и выдачи данных так называемых информационно-поисковых систем (ИПС) позволяет успешно решать многие задачи по обработке информации. Проблемы, возникающие перед разработчиками ИПС, очень сложны из-за большого количества параметров системы, разнообразных и нередко противоречивых требований пользователей, предъявляемых к ИПС, сложности и динамичности объективного мира, о котором хранятся данные. Как следствие, вопросы программного обеспечения ИПС являются весьма важными.

В середине семидесятых годов в связи с разработкой вычислительной системы коллективного пользования (ВСКП) АН Литовской ССР возникла проблема создания информационной системы и баз данных для обеспечения потребностей институтов АН Литовской ССР в средствах для хранения и обработки научно-исследовательских данных. Существенной частью работ по решению этой проблемы явилась разработка и создание математического обеспечения, что и легло в основу данной диссертации. Автором был разработан /I-5/ метод организации хранения и поиска данных с использова-

Работа выполнена в Институте математики и кибернетики АН Литовской ССР и в Лаборатории вычислительной техники и автоматизации Объединенного института ядерных исследований.

Научный руководитель:

доктор физико-математических наук
член-корреспондент АН СССР,
профессор

Николай Николаевич Говорун

Официальные оппоненты:

доктор технических наук,
профессор

Анатолий Иванович Китов

кандидат физико-математических наук

Анатолий Андреевич
Корнейчук

Ведущее предприятие:

Институт кибернетики АН УССР, г. Киев

Защита диссертации состоится "17" июня 1982 г. в "1/3" часов на заседании Ученого совета Д 047.01.04 Лаборатории вычислительной техники и автоматизации ОИЯИ г. Дубна, Московской области.

С диссертацией можно ознакомиться в библиотеке ОИЯИ.

Автореферат разослан "17" мая 1982 г.

Ученый секретарь совета

кандидат физико-математических наук

И. М. Иванченко

нием аппарата бинарных матриц и создано на его основе математическое обеспечение для ЭВМ БЭСМ-6, позволяющее осуществлять хранение, корректировку и эффективный поиск данных. Были исследованы также некоторые вопросы эффективности кодирования и декодирования информации при форматном вводе и выводе, разработаны новые алгоритмы форматного декодирования, написан вариант системных программ ввода-вывода /6,7/ для языка ФОРТРАН в мониторной системе "Дубна" /14,15/ для ЭВМ БЭСМ-6 и вариант системных программ ввода-вывода с возможностью интерактивного форматного обмена задачи с пользователем /8,9,10/.

Объемы информационных массивов ИПС непрерывно увеличиваются. Массивы обычно хранятся во внешней памяти ЭВМ, которая является сравнительно медленной, а для обработки пересылаются в быструю оперативную память ограниченного объема. Как правило, только небольшая часть хранимой информации релевантна поисковому запросу (для документальных ИПС это типичный случай). Быстрота ответа главным образом зависит от времени доступа к внешней памяти ЭВМ и количества пересылаемой для обработки информации.

Существует несколько основных методов организации информационных массивов на внешней памяти ЭВМ. Как показывают многочисленные исследования советских и зарубежных ученых (Китов А.И. /17/, Курбаков К.И. /18/, Селетков С.Н. и Волков Б.Г. /21/, Кнут Д. /19/, Cardenas A.F. /22/, Martin J. /26/, Lowe T.C. /24/, Siler K.F. /27/ и др.) основными являются списковые /16, 17, 18, 22, 23, 28/ и инверсные /17, 19, 21, 23/ методы хранения информации. При использовании списковых методов относительно легко вести поиск по простым запросам и корректировать данные, но возникают большие трудности при перестройке структуры хранимых данных и при выполнении сложных запросов по поиску /19, 22, 27/. Инверсные методы эффективны для поиска данных,

но требуют обширных объемов внешней памяти ЭВМ и создают большие трудности при вводе и коррекции данных /22, 27/. Поэтому чаще всего используются смешанные методы хранения /11, 12, 13, 19, 20, 25, 29/. При разработке любого метода организации информационных массивов надо исходить из следующих основных возможностей ускорения доступа к информации:

- 1) уменьшение объема передаваемых данных путем сжатия;
- 2) отказ от пересылки на обработку блоков данных, о которых заведомо известно, что в них нет интересующей информации;
- 3) выбор только той части данных, которая действительно необходима для обработки;
- 4) использование возможностей параллельной обработки;
- 5) улучшение параметров используемой аппаратуры.

Поэтому хранение информационных массивов во внешней памяти ЭВМ целесообразно организовывать таким образом, чтобы максимально удовлетворить противоречивые требования к быстродействию поиска, объему занимаемой внешней памяти ЭВМ, возможности работы в реальном масштабе времени. При этом следует учитывать вышеперечисленные способы ускорения доступа к данным.

Цель работы. Целью работы явилась разработка методов и программного обеспечения для хранения и поиска данных в ИПС, у которых информационные массивы большие, поиск является наиболее часто используемой процедурой и коррекция данных производится сравнительно редко, и исследование ряда вопросов эффективного обмена информацией.

Были поставлены следующие задачи:

- исследовать структуры данных и существующие методы организации информационных массивов ИПС;
- разработать новые методы организации и создать программное обеспечение ЭВМ БЭСМ-6 для ИПС с большими информационными массивами.

вами;

- исследовать некоторые вопросы эффективности форматного обмена, разработать и внедрить эффективные алгоритмы форматного декодирования.

Научная новизна работы. На основе метода бинарных матриц /29/ автором разработан модифицированный метод бинарных матриц, позволяющий работать с существенно большими массивами данных и использовать характеристики, по отношению к которым объект может обладать множеством моментных значений /1,2,4/. Используя модифицированный метод бинарных матриц, можно:

- 1) производить поиск данных с большим быстродействием;
- 2) использовать минимальный объем внешней памяти ЭВМ для размещения информационных массивов ИПС;
- 3) производить поиск по интервалу значений практически с такой же скоростью, как и для отдельных значений;
- 4) производить поиск данных по сложным запросам;
- 5) легко производить перестройку структуры данных.

Автором создано программное обеспечение на ЭВМ БЭСМ-6 для реализации модифицированного метода бинарных матриц.

Разработан метод трехуровневой организации информационных массивов ИПС, основанный на модифицированном методе бинарных матриц, удовлетворяющий требованиям, предъявляемым к ИПС, работающим в режиме реального времени. Разработаны алгоритмы ввода, корректировки и поиска информации при трехуровневой организации информационных массивов ИПС.

Проведено моделирование двухуровневой ИПС и получены экспериментальные результаты.

Разработаны новые алгоритмы декодирования при форматном обмене, позволяющие ускорить форматный вывод почти в два раза.

Практическая ценность работы. На основании модифицированно-

го метода бинарных матриц автором создано программное обеспечение для ИПС, которое используется в Институте права, философии и социологии АН Литовской ССР, Институте геологии и гидрогеологии Литовской ССР и других институтах АН Литовской ССР.

Разработанные автором алгоритмы использованы в новом варианте системных программ ввода-вывода для языка ФОРТРАН на ЭВМ БЭСМ-6. Программы включены в стандартное математическое обеспечение БЭСМ-6. Годовая экономия от использования этих программ для одной ЭВМ составляет от 60 до 100 часов времени центрального процессора БЭСМ-6.

Создан новый вариант системных программ ввода-вывода, обеспечивающий возможность интерактивного обмена задачи с терминалом. Программы используются в ВЦ АН Литовской ССР и в ЛВТА ОИЯИ для работы с терминалами, подключенными к ЭВМ БЭСМ-6 непосредственно или через концентратор.

Апробация работы. Результаты работы, положенные в основу диссертации, докладывались на Всесоюзной конференции по математической логике и программированию (Паланга, 1980), на республиканских конференциях математиков ЛитССР (Вильнюс 1975, 1977, 1979; Каунас 1978), республиканских семинарах по программированию, а также на научных семинарах ИМФ АН Литовской ССР и ЛВТА ОИЯИ.

Публикации. По материалам диссертации опубликовано 13 работ.

Объем работы. Диссертация состоит из введения, трех глав, 5 приложений, списка литературы и содержит 114 страниц машинописного текста, включая 5 рисунков, 10 таблиц и список литературы.

СОДЕРЖАНИЕ РАБОТЫ

Во введении формулируется постановка задачи на основе анализа наиболее распространенных методов хранения данных. Рассмотрены достоинства и недостатки этих методов как в отношении к

объему внешней памяти ЭВМ, требуемой для размещения информационных массивов, так и в отношении к быстродействию поиска. Обсуждаются возможности ускорения доступа к данным, приводится содержание диссертации по главам.

Первая глава посвящена вопросам организации информационных массивов ИПС. В свете проблем, возникающих при построении ИПС /3/, обсуждаются основные операции, производимые ИПС и трудности, которые возникают при реализации этих операций. Формально определяется математическая модель ИПС. Вводятся бинарные матрицы, обсуждаются вопросы хранения матриц в памяти ЭВМ, проводится сравнение с другими методами хранения по объемам памяти, требуемой для размещения данных.

Информационно-поисковой системой назовем кортеж $\lambda = \{E, C, A, D, S, F\}$ где E - множество объектов, учтенных в системе:

$$E = \{e_i, 1 \leq i \leq N\}$$

C - множество характеристик, описывающих любой объект системы:

$$C = \{c_j, 1 \leq j \leq M\}$$

A - множество множеств возможных значений для характеристик c

$$A = \{A_j, 1 \leq j \leq M\}$$

$$A_j = \{\alpha_j^k, 1 \leq j \leq M, k \in P_j\}$$

$$P_j = \{1, 2, \dots, p_j\}$$

D - область значений функции F.

S - функция двух переменных на $(D \times D)$ с областью значений $\{0, 1\}$.

Пусть множества A_j конечные. Тогда определим F следующим образом:

$$F = \{F_j, 1 \leq j \leq M, F_j: A_j \rightarrow P_j\}$$

$$F_j: \alpha_j^k \rightarrow k \text{ и } F_j^{-1}: k \rightarrow \alpha_j^k$$

Функция F взаимнооднозначно переводит множество A_j в подмножество натуральных чисел P_j . Элемент α_j^k назовем значением характеристики c_j , а число k , в которое переводится значение - кодом этого значения.

Характеристики можно подразделить на два класса в зависимости от того, может ли объект обладать в определенный момент времени одним и только одним значением данной характеристики c_j (первый класс) или иметь несколько значений (второй класс).

Так как во множестве значений A_j характеристики c_j содержатся p_j значений, то для записи любого кода k значения характеристики c_j достаточно L_j битов, где:

$$L_j = \lceil \log_2 (p_j + 1) \rceil.$$

Метод построения бинарных матриц для характеристик первого класса предложили в 1971 году Turkseu и Holzman/29/. Он состоит в том, что для каждой характеристики c_j строится матрица G_j , содержащая данные о всех объектах системы по отношению к характеристике c_j . Она состоит из N строк (i-тая строка соответствует i-тому объекту) по L_j столбцов. Каждая строка - это код значения характеристики для объекта. Если двоичное выражение кода d_i значения характеристики обозначить:

$$d_i = d_{iL_j} d_{iL_j-1} \dots d_{i1},$$

то матрица G_j будет выглядеть следующим образом:

$$G_j = \begin{matrix} d_{1L_j} & d_{1L_j-1} & \dots & d_{11} \\ \vdots & \vdots & & \vdots \\ d_{NL_j} & d_{NL_j-1} & \dots & d_{N1} \end{matrix}$$

Для характеристик второго класса матрицы G_j^i могут быть построены несколькими способами. Первый способ построения состоит в том, что в строке матрицы G_j^i помещается код значения и к нему дописывается номер объекта, обладающего этим значением /1, 2/. Альтернативным является способ, когда вместе с матрицей G_j^i , содержащей только коды значений, строится дополнительная матрица GD из Π строк по $L_n = \lceil \log_2(\Pi + 1) \rceil$ столбцов. В i -той строке матрицы GD записывается u_i - количество значений характеристики для объекта c_i . Таким образом:

$$G^i = \begin{pmatrix} d_1^i \\ \vdots \\ d_1^{u_i} \\ d_2^i \\ \vdots \\ d_2^{u_i} \\ \vdots \\ d_n^i \\ \vdots \\ d_n^{u_i} \end{pmatrix}; \quad GD = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}; \quad G^n = \begin{pmatrix} d_1^n & 1 \\ \vdots & \\ d_n^n & n \end{pmatrix}.$$

Размещение матрицы G в памяти ЭВМ происходит следующим образом. Пусть длина одного машинного слова составляет W битов. Тогда столбец G_j^k матрицы G_j^k займет $\lceil N/W \rceil$ слов. Столбцы G_j^k будут размещены в памяти ЭВМ один за другим. Аналогичным образом размещаются матрицы G^i , GD и G^n .

Для размещения матрицы G в памяти ЭВМ потребуется $\lceil L_j \cdot \lceil N/W \rceil \cdot W \rceil$ слов. Представляется интересным выяснить, какой объем памяти ЭВМ для характеристик второго класса займет дополнительная информация (матрица GD) по отношению к основной (матрица G^i). Расчеты показывают, что для объемов матриц OB_{G^i} и OB_{GD}

$$\frac{OB_{GD}}{OB_{G^i}} \approx \frac{\lceil \log_2(u+1) \rceil}{\lceil \log_2(p_j+1) \rceil \cdot \bar{u}},$$

где u - максимально допустимое и \bar{u} - среднее количество моментных значений характеристики C для объекта. Так как количество значений характеристики обычно намного больше максимального количества значений, которым может обладать один объект и, в среднем, имеется больше одного значения для объекта, то матрица GD займет лишь несколько процентов объема G^i . Возникает вопрос, в каких случаях выгоднее (с точки зрения объема памяти ЭВМ) строить матрицу G^n и в каких - G^i и GD. Оказывается, что $OB_{G^i} < OB_{G^i} + OB_{GD}$ тогда, когда выполняется неравенство $\bar{u} < \log_2(u/N)$. Сравнение с другими методами по занимаемой памяти показывает, что бинарные матрицы имеют явные преимущества перед инвертированными массивами и требуют, во всяком случае, не больших объемов памяти ЭВМ, чем при использовании списковой или мультисписковой организации.

Приведен пример преимущества использования бинарных матриц для документальных ИС.

Во второй главе обсуждены вопросы поиска при использовании модифицированного метода бинарных матриц. Формально определяется критерий поиска. Результатом поиска является вектор V в котором i -тый бит $v_i = 1$, если i -тый объект системы соответствует критерию поиска, и $v_i = 0$ в противном случае. В [29] были предложены базовые алгоритмы, использующие матрицы G первого класса и генерирующие векторы V для отношений "=", " \neq ", "<", ">", " \leq ", " \geq ".

Автором были разработаны аналогичные алгоритмы для случаев, когда матрицы G первого класса делятся на субматрицы, а также при использовании матриц G^n , G^i и GD. В качестве примера приведен базовый алгоритм поиска для отношения "равно", когда используются матрицы G^i и GD. Пусть нужно найти все объекты, которые по отношению к характеристике c_j обладают значением,

код которого в двоичном виде записывается следующим образом:

$d = d_{L-1} d_{L-2} \dots d_1$. Если обозначить через V_R -вектор длины N битов, а через V' - дополнительный вектор длины N_j , то

$$V' = \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_{N_j} \end{pmatrix} \text{ и } V' = \prod_{i=1}^L Q_i, \text{ где } Q_i = \begin{cases} G_j^i, & \text{если } k_i = 1, \\ \bar{G}_j^i, & \text{если } k_i = 0. \end{cases}$$

$$V_R = \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix}, \quad b_i = \bigcup_{t=\min}^{\max} V_t,$$

где $\min = \sum_{s=1}^{i-1} u_s + 1$, $\max = \min + u_i - 1$.

Аналогичный вид имеют другие базовые алгоритмы поиска.

Ниже приведен общий алгоритм поиска данных. Он проводится в два этапа:

1) Вводится запрос пользователя на поиск. С помощью функции F_j все значения переводятся в коды. Проверяется синтаксическая правильность оператора поиска. Если обнаружены ошибки, то работа с запросом либо прекращается (пакетный режим), либо выдается сообщение о найденных ошибках и пользователь исправляет запрос. И так до тех пор, пока он станет правильным. Оператор поиска переводится в польскую форму записи для последующей обработки;

2) Создается пустой стек векторов. Поэлементно обрабатывается польская строка запроса. Если очередной элемент строки является частью оператора поиска, относящейся к одной характерис-

тике, то с помощью базовых алгоритмов генерируется вектор и помещается в стек векторов. Если очередной элемент строки является операцией, то эта операция производится над верхним вектором стека (над двумя верхними, если операция двуместная) и результат помещается в стек векторов. После того, как вся строка обработана в нем остается лишь один вектор, являющийся результатом поиска.

Далее рассмотрены вопросы быстродействия поиска. Анализируются основные параметры, от которых зависит быстродействие, приводятся экспериментально полученные результаты. Проведены аналитические сравнения с другими наиболее известными методами поиска. Для больших ИПС предлагается трехуровневая схема хранения и алгоритмы поиска. Обсуждаются полученные экспериментальные результаты и вопросы распараллеливания предложенных автором алгоритмов поиска.

В третьей главе диссертации рассмотрены некоторые вопросы эффективности обмена информацией. Как правило, информация хранится в "сжатом" виде. Из этого, как следствие, вытекает требование, чтобы аппарат преобразования (кодирования и декодирования) информации был реализован эффективно. Особенно это относится к декодированию, поскольку однажды введенные данные используются и выдаются многократно. При использовании мониторной системы "Дубна" для ЭВМ БЭСМ-6 в программах, написанных на ФОРТРАНе для чтения и выдачи используются операторы READ, WRITE, PRINT, DECODE. Весь обмен, производимый с помощью этих операторов, осуществляется так называемой системной программой ввода-вывода. При исследовании применяемых в программе алгоритмов /15/, автором был выявлен ряд недостатков. Так при декодировании чисел типа REAL оказалось, что цифры для вывода подготавливаются порциями по 10 цифр, неэффективен подсчет порядка числа и т.п. Авто-

ром предложены новые алгоритмы и написан вариант системных программ ввода-вывода для БЭСМ-6. Применение новых алгоритмов позволило ускорить часто используемое форматное декодирование почти в два раза.

Наличие подключенных к ЭВМ терминальных устройств создает условия для эффективной отладки программ пользователей и работы в интерактивном режиме. Вместе с тем, возникает задача создания программного обеспечения, позволяющего, не меняя текста программы пользователя, производить счет как в пакетном, так и в интерактивном режимах. Описывается созданный комплекс подпрограмм для обмена текстовыми строками произвольной длины для ОС ДИСПАК /8/ и вариант системных программ ввода-вывода, позволяющий через операторы языка ФОРТРАН вести обмен как с терминалами, непосредственно подключенными к БЭСМ-6, так и с терминалами, подключенными к ней через концентратор /10/.

В заключении формулируются основные результаты, выполненные автором работы.

В приложениях приводятся некоторые алгоритмы поиска, коррекции данных и выполнения стандартных функций, используемых при генерации отчетов. Приводятся результаты моделирования ИПС.

Основные результаты работы

I. Разработка модифицированный метод бинарных матриц, позволяющий:

- производить поиск данных с большим быстродействием;
- использовать минимальный объем внешней памяти для размещения информационных массивов ИПС;
- вести поиск по интервалу значений почти с такой же скоростью, как и для отдельных значений;

- легко производить изменения в структуре хранимых данных.

Создано программное обеспечение для реализации двухуровневой ИПС, основанной на модифицируемом методе бинарных матриц на ЭВМ БЭСМ-6. Программы написаны на языке ФОРТРАН и ассемблере "Madlen"). Текст программ составляет около 8 тысяч строк (из них 1,5 тыс. на ассемблере "Madlen").

2. На основе модифицированного метода бинарных матриц разработан метод трехуровневой организации информационных массивов, удовлетворяющий требованиям, предъявляемым к ИПС, работающим в реальном масштабе времени. Разработаны алгоритмы ввода, коррекции и поиска информации при трехуровневой организации и информационных массивов.

3. Проведено моделирование двухуровневой ИПС и получены экспериментальные результаты.

4. Разработаны новые алгоритмы декодирования при форматном обмене на языке ФОРТРАН. На основе этих алгоритмов создан вариант системных программ ввода-вывода в мониторной системе "Дубна" для ЭВМ БЭСМ-6, а также вариант, позволяющий вести обмен с терминалами (как подсоединенными к БЭСМ-6 непосредственно, так и через концентратор) посредством стандартных операторов ФОРТРАН'a READ и PRINT).

Автор приносит глубокую благодарность члену-корреспонденту АН СССР, профессору Н.Н.Поворуну за руководство и постоянное внимание к работе. Глубоко признателен многим сотрудникам НИОРЭМО и ОМОЭД ЛВТА и ОПО ИМК, с которыми в течение нескольких лет довелось работать, за помощь в работе и полезные обсуждения.

ЛИТЕРАТУРА

- I. Залаторюс Ю.П. К вопросу структурной организации базы данных ИПС. В сб.: Программирование ЭВМ, вып. I, ИМК АН ЛитССР, Вильнюс, 1979, с.39-54.

2. Залаторюс Ю.П. Вопросы поиска в многоуровневых информационно поисковых системах. - В сб.: Программирование ЭВМ, вып.4, ИМК АН ЛитССР, Вильнюс, 1980, с.35-52.
3. Залаторюс Ю.П., Румшас П.Д. Некоторые вопросы построения фактографических информационно-поисковых систем. - В сб.: Программирование ЭВМ, вып.2, ИМК АН ЛитССР, Вильнюс, 1979, с.25-33.
4. Залаторюс Ю.П., Масклене Г.В., Румшас П.Д. К вопросу описания структуры данных. - В сб.: Программирование ЭВМ, вып.2, ИМК АН ЛитССР, Вильнюс, 1979, с.39-54.
5. Залаторюс Ю.П., Сицонене Д.Ю. Вопросы генерации отчетов, используя бинарные матрицы. - В сб.: Программирование ЭВМ, вып.4, ИМК АН ЛитССР, Вильнюс, 1980, с.53-60.
6. Залаторюс Ю.П. Об ускорении форматного вывода в системе "Дубна" на ЭВМ БЭСМ-6. - ОИЯИ, IO-8873, Дубна, 1975.
7. Залаторюс Ю.П. Новый вариант ввода-вывода в системе "Дубна" (Подробное описание). - ОИЯИ, БI-IO-877I, Дубна, 1975.
8. Залаторюс Ю.П. Комплекс подпрограмм для обмена текстовыми строками с терминалом. - Тезисы докладов на конф. молодых ученых ИМК АН ЛитССР и ИМ АН БССР, Вильнюс, 1978.
9. Залаторюс Ю.П. Вариант системных подпрограмм ввода-вывода для интерактивной работы пользователя с программой. - Тезисы докладов на конф. молодых ученых ИМК АН ЛитССР и ИМ АН БССР, Вильнюс, 1978.
- IO. Залаторюс Ю.П., Каданцев С.Г. Программное обеспечение БЭСМ-6 для обмена информацией между задачей пользователя и терминалом через концентратор. - ОИЯИ, PII-80-290, Дубна, 1980.
- II. Говорун Н.Н., Гусев А.В., Заикин Н.С., Залаторюс Ю.П. и др. Программное обеспечение системы "Кадры". - ОИЯИ, IO-IO95I, Дубна, 1977.

12. Говорун Н.Н., Залаторюс Ю.П. и др. Организация информационных массивов системы "Кадры". - ОИЯИ, IO-II05I, Дубна, 1977.
13. Говорун Н.Н., Залаторюс Ю.П. и др. Язык входных сообщений и язык запроса информационной системы "Кадры". - ОИЯИ, IO-II052, Дубна, 1977.
14. Говорун Н.Н. и др. Мониторная система "Дубна" для ЭВМ БЭСМ-6. - В сб.: Труды 2-ой Всесоюзной конференции по программированию, вып. Ж. ВЦ СО АН СССР, Новосибирск, 1970, с.5-24.
15. Загинайко В.А. Реализация обработки операторов обмена с внешними устройствами языка ФОРТРАН на машине БЭСМ-6. - БИ-II-3907, Дубна, 1968.
16. Андон Ф.И., Стогний А.А. Система ИНФОР общее описание. - УСИМ № 4, 1974. - 26 с.
17. Китов А.И. Программирование экономических и управленческих задач. - М.: Советское радио, 1971.
18. Курбаков К.И. Информационно-логические системы. - М.: Знание, 1967.
19. Кнут Д. Искусство программирования для ЭВМ. Т.3. - М.: Мир, 1978.
20. Руденко Р.В., Элланская Л.В. и др. Опыт использования СУБД ОКА в разработке подсистемы конструкторско-технологической подготовки производства. - УСИМ, № 4, 1981. - 49 с.
21. Селетков С.Н., Волков В.Г. Хранение и поиск данных в ЭВМ. - М.: Советское радио, 1971.
22. Cardenas A.F. Evaluation and selection of file organization - a model and a system. - SACM, vol.16, N 9, (Sept., 1973).
23. Inglis J. Inverted indexes and multilist structures. - The CJ, vol. 17, N 1, 1974.

24. Lowe T.C. The influence of data base characteristics on direct access file organisation. - J.ACM, v.I4, N 4, (Oct., 1968).
25. Ium V.Y. Multi-attribute retrieval with combined indexes. - CACM, vol. 13, N II, (Nov., 1970).
26. Martin J. Computer data base organisation. - Prentice-hall. INC, Englewood Clifs, N.J., 1978.
27. Siler K.F. A stochastic evaluation model for database organisations in data retrieval systems. - CACM, vol. 19, N 2, (Feb., 1976).
28. Sussenguth E.N. The use of tree structures for processing files. - CACM, vol. 6, N 5, (May 1963).
29. Turksen L.B., Holzman A.G. Binary conversion and related design concepts for management data systems. - INFOR, vol.9, N 2 (July, 1971).