

284/83



сообщения
объединенного
института
ядерных
исследований
дубна

10/1-83

10-82-686

А.М.Ершов

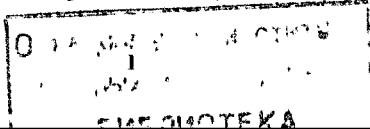
ДИНАМИЧЕСКОЕ ПОСТРОЕНИЕ АРГУМЕНТОВ
ПОИСКА СЕГМЕНТОВ В СИСТЕМЕ УПРАВЛЕНИЯ
БАЗАМИ ДАННЫХ "ОКА"
НА ОСНОВАНИИ ЗАПРОСОВ ПОЛЬЗОВАТЕЛЕЙ

1982

ВВЕДЕНИЕ

С каждым годом неуклонно возрастает роль информационного обслуживания самых различных сторон человеческой деятельности: научной, технической, экономической. В связи с этим особое внимание уделяется создаваемым на базе ЭВМ информационным системам (ИС). Накапливаемая в ИС информация представляет собой модель некоторой области реального мира. Поэтому, конструируя ИС, необходимо четко представлять, какой предметной области и в какой мере адекватна содержащаяся в ней информация. Понятие адекватности как общего явления не формализовано. Однако в каждом конкретном случае, как правило, известен набор требований, предъявляемых к информационной модели, и на практике обычно делается допущение о структуре моделируемой области. В настоящее время используется ряд структур моделируемой области: дескрипторная, объектно-характеристическая, триадная, иерархическая, сетевая, релятивная^{1/}. Во многих случаях, когда в качестве допущения можно принять гипотезу об иерархической структуре описываемой области реального мира, для реализации ИС на машинах ЕС ЭВМ может быть использована система управления базами данных (СУБД) "ОКА"^{2/}.

В работе рассмотрен метод разработки в среде СУБД "ОКА" прикладного программного обеспечения для создания информационно-поисковой системы фактографического типа, предназначенной для поиска, отбора и извлечения информации из баз данных по запросам пользователей. Основной акцент делается на обеспечение независимости прикладных программ от изменений в составе и характеристиках элементов описаний информационных объектов, называемых также поисковыми признаками, или поисковыми реквизитами. Независимость программ от физических структур хранения данных обеспечивается СУБД. Описан метод динамического построения аргумента поиска сегмента, реализующий обработку сложных запросов. Сложным называется запрос, состоящий из нескольких элементар-



ных высказываний, соединенных между собой знаками конъюнкции и дизъюнкции^{3,4/}.

Интерфейс прикладной программы в СУБД "ОКА"

Возможность обработки баз данных в системе "ОКА" реализуется при помощи языка BETA, использующего соответствующие блоки связи с программой (PCB - Program Communication Block). Язык BETA обеспечивает независимость прикладных программ от методов доступа, от физической организации памяти и от тех устройств, на которых хранятся нужные данные. PCB представляет собой таблицу, через которую осуществляется связь прикладной программы с определенной базой данных. Все PCB хранятся отдельно от прикладных программ и во время работы используются языком BETA. Средством, с помощью которого язык BETA передает данные прикладной программе, является стандартный оператор вызова языка высокого уровня. Формат оператора вызова для языка КОБОЛ выглядит следующим образом:

```
CALL 'CBLTBET' USING FUNC,PCB,WORK-AREA,SSA1,... SSAN
```

где FUNC - имя четырехсимвольного поля, описывающего запрашиваемую функцию ввода-вывода вызова языка BETA;
PCB - адрес указателя на имя блока связи с программой для базы данных;
WORK-AREA - адрес рабочей области ввода-вывода;
SSA1-SSAN - адреса аргументов поиска сегментов.

С точки зрения прикладной программы логическая структура базы данных представляет собой иерархическую структуру сегментов. Сегмент - это элемент данных фиксированной длины, содержащий одно или несколько логически взаимосвязанных полей данных. Сегмент является наименьшей информационной единицей обмена между прикладной программой и языком BETA.

Аргумент поиска сегмента (SSA-Search Segment Argument) включает в себе всю необходимую информацию, связанную с заданием на поиск в базе данных и используемую языком BETA.

Задание критериев поиска в базе данных

При поиске в базе данных язык BETA использует заданные в SSA предложения квалификации для того, чтобы определить, удовлетворяет ли сегмент спецификациям пользователя. Под предложением квалификации, или квалификационным предложением, понимают условное отношение, указанное для отдельного поля сегмента. На предмет выполнения этого условия будет просматриваться база данных при поиске. В SSA можно указать до восьми предложений квалификации. Структура аргумента поиска сегмента выглядит следующим образом:

имя (квалификационное + квалификационное)
сегмента предложение I ж ... предложение 8

Для ясности здесь опущены коды команд, которые могут указываться дополнительно для спецификации специальных режимов обработки. Символы + и ж используются для обозначения логических операторов конъюнкции и дизъюнкции. В свою очередь, квалификационное предложение состоит из имени поля сегмента, оператора отношения и значения для сравнения.

В обычной ситуации при создании прикладных программ известен весь набор запросов, связанных с обработкой баз данных. Поэтому, как правило, в программе резервируются рабочие поля, в которых размещаются все необходимые SSA. При использовании этих полей в SSA только представляются необходимые значения для сравнения. Однако при создании информационной системы с большим количеством поисковых признаков заранее не известны аргументы поиска сегментов, так как практически невозможно учесть все возможные комбинации квалификационных предложений, необходимых для обработки сложных запросов. При этом потребовалось бы резервирование огромного количества рабочих полей для SSA, а прикладная программа стала бы чрезвычайно громоздкой и, что самое важное, чувствительной к изменениям в составе и форматах поисковых признаков, а также к изменениям в логической структуре используемых баз данных. Для устранения перечисленных недостатков необходимо реализовать динамическое построение аргументов поиска сегментов на основании информации, получаемой из входного запроса.

Использование справочной базы данных

Для обеспечения независимости прикладной программы от состава и характеристик поисковых реквизитов необходимо отдельное хранение всей определяющей их информации. Для организации такого хранения используется специальная справочная база данных (см. рис.1). Основная база данных, содержащая описания информационных объектов, называется информационной. Для идентификации указанных в запросе поисковых признаков каждому признаку ставится в соответствие отдельное поле одного из сегментов информационной базы данных. Это соответствие отражено в справочной базе данных. Логическая структура справочной базы данных имеет два уровня и состоит из двух типов сегментов. Корневой сегмент содержит следующую информацию о поисковых реквизитах:

- номер или идентификатор поискового реквизита, присвоенный ему в информационной системе;
- наименование поискового реквизита, используемое при формировании выходных отчетов и справок, выдаваемых в ответ на запрос;

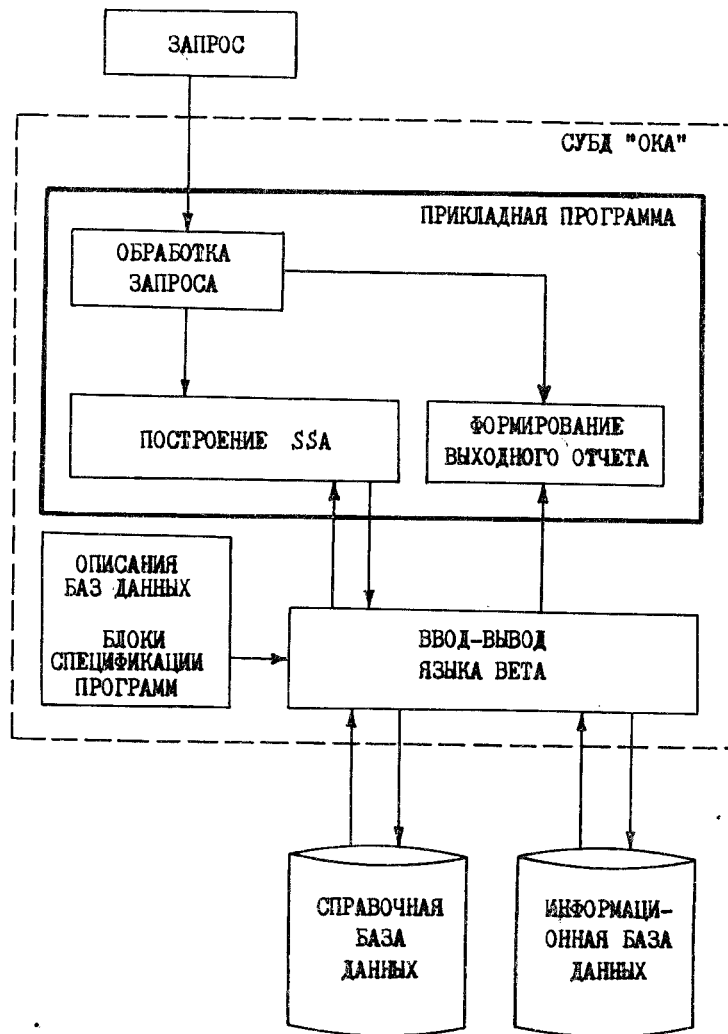


Рис. 1. Функционирование информационной системы с динамическим построением SSA.

- имя сегмента, содержащего значение данного поискового реквизита, в информационной базе данных;
- имя поля, соответствующего значению данного поискового реквизита, и его начальная позиция в сегменте информационной базы данных;
- длина значения реквизита, необходимая при формировании SSA ;
- флаг, указывающий, является ли значение данного поискового реквизита каждый раз уникальным или всегда представляет собой элемент определенного фиксированного набора значений. В последнем случае каждый элемент набора значений вместе с его ключевым кодом содержится в справочной базе данных в виде реализации подчиненного сегмента;
- тип реквизита, определяющий посредством кода его внутреннее представление.

Так как обычно значительная часть поисковых признаков в ИС имеет фиксированный набор значений (списки подразделений, должностей и т.п.), применение справочной базы данных снижает процент дублирования информации и позволяет эффективно использовать внешнюю память для размещения информационной базы данных, в которой при такой организации находятся только соответствующие ключевые коды значений.

Обращаясь к справочной базе данных, прикладная программа получает всю необходимую информацию: имена полей, их расположение в сегменте и т.п. Поэтому при изменении состава или характеристик поисковых реквизитов потребуется только корректировка справочной базы данных.

Необходимо отметить, что некоторые характеристики полей, сегментов и самой информационной базы данных могут быть получены при помощи непосредственного обращения к ее описанию в библиотеке описаний баз данных системы "ОКА". Однако при этом невозможно получить следующие необходимые сведения из числа рассмотренных выше: информацию о соответствии между поисковыми реквизитами и именами полей сегментов, наименования и тип поисковых реквизитов, принадлежность набора значений. Вся эта информация заключена в справочной базе данных.

Динамическое построение аргумента поиска сегмента

Для динамического построения SSA в прикладной программе требуется зарезервировать несколько рабочих полей, используемых затем для формирования в них SSA. Их количество равно числу уровней в логической иерархии информационной базы данных. Длина каждого поля равна:

$$\text{CONST} + (\text{VALUE1} + \text{VALUE2} + \dots + \text{VALUE8}) .$$

Постоянная CONST определяет количество позиций, необходимых для размещения в соответствии со структурой SSA имен сегмента и полей, а также служебных знаков SSA для восьми квалификационных предложений (круглых скобок, символов логических операторов и операторов отношений): CONST = 97. Переменные VALUE1 - VALUE8 представляют собой восемь наибольших длин значений поисковых реквизитов соответствующего иерархического уровня информационной базы данных.

В каждом конкретном случае реализация языка запросов может быть различной. Важно, чтобы он предоставлял возможность формировать сложные запросы, состоящие из нескольких элементарных высказываний, объединенных знаками конъюнкции и дизъюнкции. Усложнение языка запросов влечет за собой соответствующее усложнение той части программного обеспечения, которая отвечает за разбор запроса. Поисковые реквизиты могут задаваться при помощи номеров, идентификаторов и т.д. Результатом обработки запроса пользователя ИС (см. рис.1), в процессе которой выполняется синтаксический разбор введенных предложений, являются сведения о заданных логических взаимоотношениях между отдельными элементами запроса, о поисковых реквизитах и критериях отбора информации из баз данных - операторах отношений и значениях для сравнения. Вся эта информация, а также сведения из справочной базы данных служат в качестве исходных при динамическом построении аргументов поиска сегментов (см. рис.2). Формирование SSA производится программой в зарезервированных для этой цели рабочих полях в соответствии со структурой SSA. При построении каждого квалификационного предложения особое внимание уделяется заполнению полей, содержащих значения для сравнения. При этом требуется учитывать выравнивание значений и тот факт, является ли соответствующий поисковый признак буквенно-цифровым или числовым, если числовым, то каково его внутреннее представление. Необходимую для этого информацию содержат поля "длина значения реквизита" и "тип реквизита" корневого сегмента справочной базы данных. Сформированные SSA используются в вызове при обращении к языку BEGA для поиска и извлечения информации из баз данных.

Заключение и пример реализации

В простейшем случае, когда логическая структура информационной базы данных содержит один сегмент, в справочной базе данных хранится минимум информации: имена полей сегмента и связанные с ними характеристики. В более сложном случае в справочной базе данных содержится также информация о сегментах и, возможно, о логической структуре информационной базы данных. Для описания логической структуры достаточно указания номера иерархического уровня для каждого сегмента.

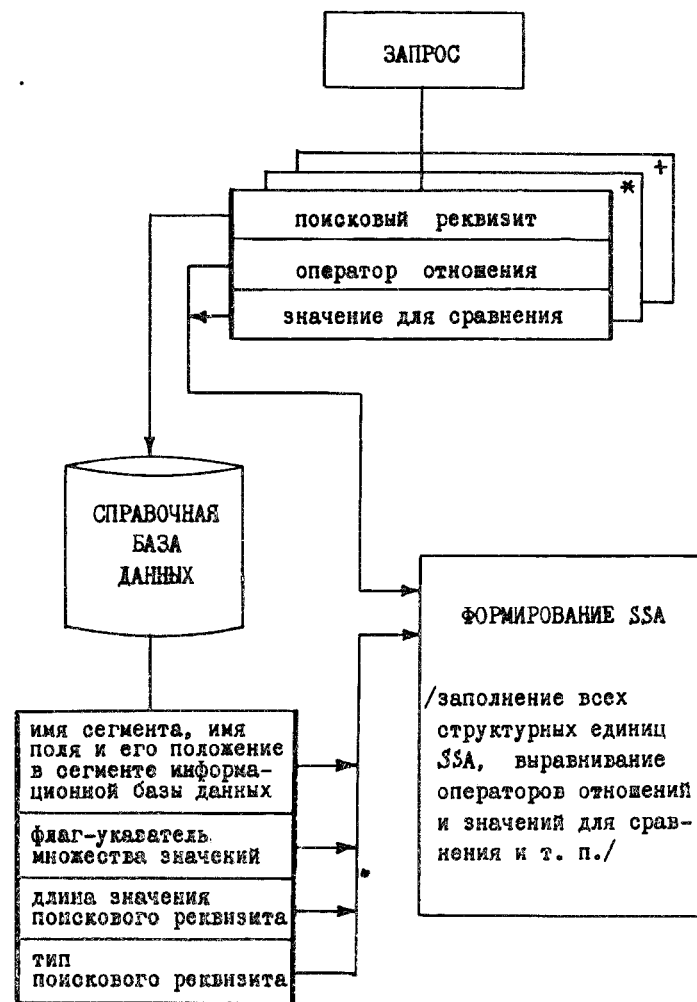


Рис. 2. Динамическое построение SSA.

Так как каждому поисковому признаку соответствует отдельное поле некоторого сегмента в информационной базе данных, необходимо рассмотреть связанные с этим ограничения системы "ОКА". СУБД "ОКА" допускает определение до 255 полей в одном сегменте и до 1000 полей в одной базе данных. Поэтому, если в создаваемой информационной системе имеется более 1000 поисковых признаков, потребуется распределение обрабатываемой информации между двумя или более информационными базами данных. В случае использования нескольких информационных баз данных в справочных сведениях по каждому поисковому реквизиту дополнительно указывается содержащая его информационная база данных. Соответственно в прикладной программе увеличивается число вызовов для обращения к языку ВЕГА.

Ограничение, связанное с тем, что в аргументе поиска сегмента должно присутствовать не более восьми предложений квалификации, то есть в запросе не должно находиться больше восьми поисковых признаков, относящихся к одному типу сегмента в информационной базе данных, не является существенно суживающим возможности обработки. В большинстве случаев поисковые реквизиты рассредоточены по сегментам различных типов и, как правило, указанное ограничение не играет роли.

При работе СУБД "ОКА" в пакетном режиме запросы на поиск в ИС принимаются в виде входной информации соответствующей прикладной программы. При работе в режиме телеобработки запросы целесообразно реализовать в качестве входных сообщений (транзакций) системы "ОКА".

Независимость прикладных программ от используемых данных, обеспечиваемая СУБД, дает возможность в каждом конкретном случае выбрать оптимальную физическую структуру хранения баз данных в зависимости от того, какой фактор является определяющим: время поиска или расходная внешняя память^{5/}. Например, в режиме телеобработки желательнее малое время ответа ИС, что в пакетном режиме не так важно. Необходимо отметить, что если в СУБД не реализовано вторичное индексирование^{6/}, позволяющее вести индекс одновременно по нескольким полям сегмента, запросы, не затрагивающие непосредственно ключевых полей сегментов, приводят к просмотру всей базы данных. Учитывая этот факт, применение рассмотренного метода можно рекомендовать для ИС, имеющих в основе базы данных среднего объема.

Динамическое построение аргументов поиска сегментов требует создания справочной базы данных, а также некоторых дополнительных затрат времени непосредственно на формирование SSA при обработке запроса. Однако эти затраты полностью окупаются, поскольку программное обеспечение, которое при таком подходе приобретает простоту и ясность, становится независимым от состава и характеристик используемых в ИС признаков и от изменений в описаниях информационных объектов. Кроме того,

появляется возможность использовать весь комплекс служебных средств, предоставляемых системой "ОКА": реорганизацию и восстановление баз данных, ведение системного журнала и т.д.

На основе метода динамического построения аргумента поиска сегмента на ЭВМ ЕС-1060 ОИЯИ в среде СУБД "ОКА" создан комплекс программ, представляющий собой модель ИС "ИСКРА" (информационная система для комсомольской работы). В качестве примера приводится образец запроса в этой информационно-поисковой системе. Запрос - найти фамилии комсомольцев с окладом или среднемесячной зарплатой меньше 150 руб., имеющих больше одного ребенка и проживающих не в отдельной квартире; выдать на печать сведения об образовании, должности, жилищных условиях, количестве жилплощади на человека, числе детей и номер рабочего телефона:

```
FIND 7=2 * 23 <150 * 28> I * 24]=I
```

```
PRINT 9, I4, 24, 25, 28, 30
```

Для идентификации поисковых признаков в запросе используются их номера:

- 7 - партийность,
- 9 - образование,
- I4 - должность,
- 23 - оклад или среднемесячная зарплата,
- 24 - жилищные условия,
- 25 - количество жилплощади на человека;
- 28 - число детей,
- 30 - рабочий телефон.

Язык запросов ориентирован на форму анкеты, которая используется при сборе исходных данных для ИС. Оператор **FIND** содержит сведения о критериях отбора информации в информационной базе данных. В зависимости от вида поискового признака его значение задается либо соответствующим кодом (номером) из некоторого фиксированного набора, либо непосредственно. Например, 7=2 в запросе означает "партийность - член ВЛКСМ" (код - 2), а 24]=I - "жилищные условия - не отдельная квартира" (код - I). Оператор **PRINT** содержит сведения о поисковых признаках, соответствующих элементам описаний, выдаваемым на печать в выходном отчете.

ЛИТЕРАТУРА

1. Криницкий Н.А., Миронов Г.А., Фролов Г.Д. Автоматизированные информационные системы. "Наука", М., 1982.
2. Андон Ф.И., Довгопольный В.Г. и др. Основные положения системы управления базами данных "ОКА". "Управляющие системы и машины", 1977, № 2.
3. Макаровский Б.Н. Информационные системы и структуры данных. "Статистика", М., 1980.
4. Валентинов В.В. Информационно-поисковая система в АСУ. "Статистика", М., 1976.
5. Ершов А.М., Сычев П.П. ОИЯИ, IO-82-464, Дубна, 1982.
6. Дейт К. Введение в системы баз данных. "Наука", М., 1980.

Рукопись поступила в издательский отдел
23 сентября 1982 года.

НЕТ ЛИ ПРОБЕЛОВ В ВАШЕЙ БИБЛИОТЕКЕ?

Вы можете получить по почте перечисленные ниже книги, если они не были заказаны ранее.

D13-11182	Труды IX Международного симпозиума по ядерной электронике. Варна, 1977.	5 р. 00 к.
D17-11490	Труды Международного симпозиума по избранным проблемам статистической механики. Дубна, 1977.	6 р. 00 к.
D6-11574	Сборник аннотаций XV совещания по ядерной спектроскопии и теории ядра. Дубна, 1978.	2 р. 50 к.
D3-11787	Труды III Международной школы по нейтронной физике. Алушта, 1978.	3 р. 00 к.
D13-11807	Труды III Международного совещания по пропорциональным и дрейфовым камерам. Дубна, 1978.	6 р. 00 к.
D1,2-12036	Труды VI Всесоюзного совещания по ускорителям заряженных частиц. Дубна, 1978 /2 тома/	7 р. 40 к.
D1,2-12450	Труды V Международного семинара по проблемам физики высоких энергий. Дубна, 1978	5 р. 00 к.
D1,2-12450	Труды XII Международной школы молодых ученых по физике высоких энергий. Приморско, НРБ, 1978.	3 р. 00 к.
D11-80-13	Труды VII Всесоюзного совещания по ускорителям заряженных частиц, Дубна, 1980 /2 тома/	8 р. 00 к.
D4-80-271	Труды рабочего совещания по системам и методам аналитических вычислений на ЭВМ и их применению в теоретической физике, Дубна, 1979	3 р. 50 к.
D4-80-385	Труды Международной конференции по проблемам нескольких тел в ядерной физике. Дубна, 1979.	3 р. 00 к.
D2-81-543	Труды Международной школы по структуре ядра. Алушта, 1980.	5 р. 00 к.
D2-81-543	Труды VI Международного совещания по проблемам квантовой теории поля. Алушта, 1981	2 р. 50 к.
D10,11-81-622	Труды VII Международного совещания по проблемам математического моделирования в ядерно-физических исследованиях. Дубна, 1980	2 р. 50 к.
D1,2-81-728	Труды VI Международного семинара по проблемам физики высоких энергий. Дубна, 1981.	3 р. 60 к.
D17-81-758	Труды II Международного симпозиума по избранным проблемам статистической механики. Дубна, 1981.	5 р. 40 к.
D1,2-82-27	Труды Международного симпозиума по поляризационным явлениям в физике высоких энергий. Дубна, 1981.	3 р. 20 к.
P18-82-117	Труды IV совещания по использованию новых ядерно-физических методов для решения научно-технических и народнохозяйственных задач. Дубна, 1981.	3 р. 80 к.

Заказы на упомянутые книги могут быть направлены по адресу:
101000 Москва, Главпочтамт, п/я 79
Издательский отдел Объединенного института ядерных исследований

**ТЕМАТИЧЕСКИЕ КАТЕГОРИИ ПУБЛИКАЦИЙ
ОБЪЕДИНЕННОГО ИНСТИТУТА ЯДЕРНЫХ
ИССЛЕДОВАНИЙ**

Индекс	Тематика
1.	Экспериментальная физика высоких энергий
2.	Теоретическая физика высоких энергий
3.	Экспериментальная нейтронная физика
4.	Теоретическая физика низких энергий
5.	Математика
6.	Ядерная спектроскопия и радиохимия
7.	Физика тяжелых ионов
8.	Криогеника
9.	Ускорители
10.	Автоматизация обработки экспериментальных данных
11.	Вычислительная математика и техника
12.	Химия
13.	Техника физического эксперимента
14.	Исследования твердых тел и жидкостей ядерными методами
15.	Экспериментальная физика ядерных реакций при низких энергиях
16.	Дозиметрия и физика защиты
17.	Теория конденсированного состояния
18.	Использование результатов и методов фундаментальных физических исследований в смежных областях науки и техники
19.	Биофизика

Ершов А.М.

10-82-686

Динамическое построение аргументов поиска сегментов в системе управления базами данных "Ока" на основании запросов пользователей

Рассматривается возможность использования системы управления базами данных /СУБД/ "Ока" для создания информационно-поисковой системы. Предлагается метод динамического построения аргумента поиска сегмента для обработки запросов пользователей. Описывается реализация метода при создании в среде СУБД "Ока" модели информационно-поисковой системы со сложными запросами.

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1982

Ershov A.M.

10-82-686

Dynamic Construction of Search Segment Arguments in "Oka" Data Base Management System on the Base of Users' Requests

The possibility to use "Oka" data base management system /DBMS/ for creating the information retrieval system is considered. The method of dynamic construction of search segment argument for user requests processing is presented. The realization of the method for a model of information retrieval system with compound requests under "Oka" DBMS is described.

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

Communication of the Joint Institute for Nuclear Research. Dubna 1982

Перевод О.С.Виноградовой.