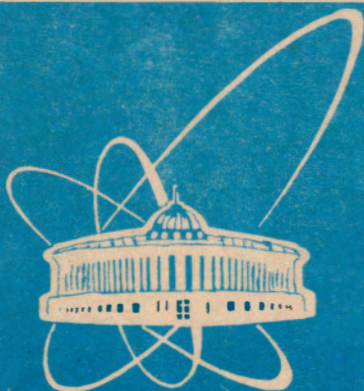


94-395



сообщения
объединенного
института
ядерных
исследований
дубна

P11-94-395

А.М.Ершов¹

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ
БИБЛИОГРАФИЧЕСКОЙ
ИНФОРМАЦИОННО-ПОИСКОВОЙ
СИСТЕМЫ BLESC
НАУЧНО-ТЕХНИЧЕСКОЙ БИБЛИОТЕКИ ОИЯИ

¹E-mail: ershov@lcta10.jinr.dubna.su

1. ВВЕДЕНИЕ

Среди широкого спектра работ, выполняемых в ЛВТА в рамках проекта развития информационной инфраструктуры ОИЯИ, одним из основных направлений является современная компьютерная поддержка научных исследований. В качестве важного элемента решения этой задачи выступает создание центра научно-технической информации на базе ряда автоматизированных информационных систем в распределенной вычислительной среде, обеспечивающих хранение, обработку и оперативное получение различного рода сведений – от библиографических указателей до фрагментов баз данных сложнейших экспериментов.

Известно, какую огромную часть времени при проведении научных исследований занимает библиографический поиск необходимой литературы и регулярный просмотр новых публикаций в исследуемой предметной области. Разработанная в ЛВТА библиографическая система BLESC предоставляет сотрудникам ОИЯИ возможность избежать утомительного изучения множества библиографических материалов в научно-технической библиотеке и в течение нескольких минут получить ссылки на интересующие публикации. И, что важно, все это – непосредственно со своего рабочего места.

2. ИНФОРМАЦИОННО-ПОИСКОВАЯ СИСТЕМА BLESC

Многим сотрудникам, активно использующим в своей повседневной работе материалы НТБ ОИЯИ, уже знакома библиографическая информационно-поисковая система НТБ. В течение 1993 года эта система функционировала в локальном варианте на отдельном персональном компьютере (ПК) типа IBM PC, установленном в зале иностранной литературы в корпусе ЛТФ. Год ее достаточно интенсивной эксплуатации (в среднем около 270 поисковых запросов в день) позволил учесть множество полезных предложений и замечаний, обеспечить высокую надежность и устойчивость работы. Создание в 1994 году локальной подсети НТБ ОИЯИ, установка специального файл-сервера и подключение его к вычислительной сети ОИЯИ позволили решить основную задачу, которая с самого начала стояла перед разработчиками: обеспечение удаленного коллективного доступа к библиографической

информации НТБ с любого ПК, подключенного к вычислительной сети ОИЯИ ETHERNET.

В текущей версии 5.07 библиографической системы BLESC реализован оперативный доступ к следующим базам данных: BOOKS – книги, полученные библиотекой, начиная с 1990 года; JINR-WORK – публикации сотрудников ОИЯИ, начиная с 1988 года; PREP-OLD и PREP-NEW – препринты и статьи, полученные соответственно до и после 1993 года. Последние две базы данных разделены, в первую очередь, исходя из соображений удобства их сопровождения. Содержимое информационной базы системы регулярно пополняется библиографическими данными о новых поступлениях в НТБ. В течение длительного периода времени эти сведения также издаются в печатном виде в качестве еженедельных экспресс-бюллетеней ОИЯИ "Статьи", "Препринты" и "Книги".

В настоящий момент в НТБ ОИЯИ нет ни технической, ни организационной возможности для хранения и обработки полнотекстовых документов. Соответственно и в системе BLESC содержатся только все основные и ключевые реквизиты публикаций. Результаты обработки поисковых запросов выдаются на экран компьютера в формате, близком к общепринятому формату оформления библиографического списка работ, что является привычным и удобным для восприятия информации.

3. ИНФОРМАЦИОННАЯ БАЗА ДАННЫХ СИСТЕМЫ BLESC

Организация многопользовательской интерактивной работы в сетевой среде предъявляет жесткие требования к создаваемому программному обеспечению. В качестве основной платформы в информационно-поисковой системе НТБ BLESC использована система управления базами данных Micro CDS/ISIS версии 3.0 (сетевой вариант), распространяемая ЮНЕСКО на некоммерческой основе среди предприятий науки, культуры и образования. Официальным дистрибьютором CDS/ISIS/М ЮНЕСКО в России является Международный центр научной и технической информации [1].

Средства CDS/ISIS ориентированы на хранение и обработку структурированных нечисловых баз данных с полями и записями переменной длины. При разработке библиографической системы BLESC был

задействован только наиболее удачный элемент CDS/ISIS – аппарат построения инвертированных списков и организации поискового процессора документального типа. Пользовательский же интерфейс был написан с использованием специализированного языка Pascal-M, включающего в себя функции непосредственного взаимодействия с базами данных CDS/ISIS.

Вся информация в базах данных НТБ ОИЯИ содержится только в текстовой форме, поэтому для представления символов, отсутствующих в обычной (альтернативной) кодировке ASCII и требующих применения некоторых графических возможностей, на самой ранней стадии был разработан несложный язык преобразований. Среди таких символов, в первую очередь, можно выделить показатели степени и индексы в формулах, греческие буквы и т.п. Как правило, отмеченные преобразования встречаются только в заглавиях публикаций. Для лучшего их понимания приведем несколько примеров. Так, для выделения показателя степени используется функция $\&\text{sup}()$, а для нижнего индекса – функция $\&\text{sub}()$. Символ "звездочка" применяется как для указания числового показателя степени, так и для выражения многих специальных символов и букв других алфавитов. Например, комбинация *a соответствует α , *b — β , *g — γ . Еще несколько выражений и их представления в оригинале: $\text{La}\&\text{sub}(2)\text{CuO}\&\text{sub}(4)$ — La_2CuO_4 ; Fe^*5^*7 — Fe^{57} ; $*\text{SE}\&\text{sub}(*g)$ — $\sum E_\gamma$. Полную справку об используемых преобразованиях неграфических символов можно получить в научно-технической библиотеке ОИЯИ.

4. ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ СИСТЕМЫ BLESC

В основном меню системы BLESC пользователю предлагается выбор следующих режимов работы:

- Выполнение поисковых запросов
- Сохранение результатов поиска
- Просмотр поисковых терминов
- Справочный просмотр рубрикатора
- Заказ литературы на абонементе
- Пополнение библиографических данных
- Выбор библиографической базы данных

Вся основная работа по библиографическому поиску, нахождению и отбору сведений об интересующих публикациях осуществляется в режиме *Выполнение поисковых запросов*. С одной стороны, для работы с системой BLESC вполне достаточно знания самых элементарных принципов построения запросов на поиск. С другой стороны, более профессиональное использование синтаксических правил, которым необходимо следовать при формулировании поисковых запросов, обеспечивает получение требуемой информации с наименьшими затратами времени и сил, поэтому следующий раздел целиком посвящен подробному описанию языка запросов.

В процессе обработки поискового запроса пользователя на экран компьютера в режиме реального времени выдается оперативный протокол выполнения операции поиска с итоговым указанием количества найденных релевантных публикаций. После того, как поиск завершен, сразу же может быть выполнен просмотр отобранных библиографических записей.

Результаты любого предшествующего успешного поиска могут быть записаны на жесткий диск в режиме *Сохранение результатов поиска*. Под успешным понимается запрос на поиск, которому удовлетворила хотя бы одна библиографическая запись из базы данных. Пользователю предлагается подтвердить имя создаваемого на диске файла, состоящее из наименования текущей поисковой базы данных, порядкового номера запроса и расширения 'ntb'. Для идентификации записанных на диск результатов поиска в созданном файле в качестве заголовка присутствуют текст введенного запроса, количество найденных по нему публикаций, дата и время выполнения поискового запроса.

Для получения оперативной справки о ключевых словах, которые могут быть заданы в запросе, предназначен режим работы *Просмотр поисковых терминов*. Несмотря на то, что этот режим носит вспомогательный характер, он оказывается исключительно полезным для ориентации пользователя в процессе составления и ввода поискового выражения. При этом на экран компьютера в алфавитном порядке по колонно выводятся все поисковые термины, принадлежащие активной базе данных. Так как этих терминов имеется огромное количество, то для установки в нужной позиции и вывода на экран требуемого диапазона списка предлагается ввести несколько начальных символов термина.

Для нахождения интересующих публикаций в качестве поисковых критериев могут быть задействованы практически любые их библиографические реквизиты. Как правило, каждого читателя интересует литература в определенной, близкой ему предметной области. Поэтому часто оказывается очень полезным и более удобным произвести поиск по конкретным предметным рубрикам. Именно для подбора нужных кодов предметных рубрик с целью их последующего использования в запросах предусматривается ввод режима *Справочный просмотр рубрикатора*.

Режим работы *Заказ литературы на абонементе* реализует информационную связь библиографической поисковой системы BLESC с разрабатываемым программным обеспечением абонента научно-технической библиотеки ОИЯИ. Этот режим предоставляет пользователю возможность для заказа найденной и интересующей его литературы непосредственно со своего рабочего места. В качестве заказываемого списка литературы используется результат последнего успешно выполненного поиска. Естественно, это требует более тщательной подготовки вводимого запроса, например, с использованием средств пошагового уточнения поискового запроса (подраздел 5.2). Предварительно список отобранной литературы для контроля высвечивается на экране, затем требуется его подтверждение и ввод номера читательского билета. Необходимо четко представлять, что данный режим не означает автоматический заказ литературы: в процессе его выполнения производится только передача сформированного заказа сотрудникам абонента НТБ.

Первоначально после инициализации рабочего сеанса связи пользователю по умолчанию предлагается для поиска база данных препринтов и статей, полученных научно-технической библиотекой в 1993 году и позднее (PREP-NEW). При необходимости выполнить поиск среди других публикаций используется режим *Выбор библиографической базы данных*. Наименование текущей, то есть активной и используемой для поиска в данный момент, базы данных всегда присутствует в правой верхней части экрана компьютера.

Пополнение библиографических данных выполняется сотрудниками НТБ ОИЯИ, ответственными за обработку и сопровождение соответствующего раздела библиотечного информационного фонда. Этот режим доступен только при предъявлении системного пароля.

5. ЯЗЫК ЗАПРОСОВ ПОИСКОВОЙ СИСТЕМЫ BLESC

В основе языка запросов библиографической информационно-поисковой системы BLESC лежит адаптированная версия оригинального поискового языка CDS/ISIS/M. Простота и доступность синтаксических языковых конструкций не требует специальных знаний при формировании запросов и обеспечивает высокую эффективность общения с системой даже неподготовленного пользователя.

С целью иллюстрации поисковых возможностей системы BLESC в подразделах 5.1 и 5.2 подробно на конкретных примерах рассматриваются простые и сложные варианты запросов к библиографическим базам данных. Приводимые запросы играют также вспомогательную роль и могут быть использованы в качестве образца при самостоятельной практической работе с системой.

При формулировке поискового выражения используются поисковые термины, называемые также терминами доступа. В качестве них могут выступать слова заглавия публикации, фамилии авторов, шифры, коды, ключевые слова предметной области и другие элементы библиографической записи. Например, введенное слово `graph` уже может служить запросом на поиск, в процессе выполнения которого из базы данных будут отобраны все публикации, содержащие в своих описательных реквизитах слово `graph`. Достоинством системы BLESC является независимость результатов поиска от того, в каком регистре (строчными или прописными буквами) был введен запрос. Перед выполнением операции поиска производится перевод всех строчных символов в их прописные аналоги; поэтому будут эквивалентны поисковые запросы, состоящие из терминов: `graph`, `Graph`, `GRAPH` и т.д.

Помимо поиска на точное соответствие часто используется более полезный и удобный поиск слов или терминов, усеченных справа. Усечение справа указывается размещением знака доллара (\$) сразу же за последним символом поисковой основы. Например, по запросу `graph$` в базе данных будут найдены все публикации, содержащие слова, начинающиеся с символов "graph" (`graphical`, `graphics`, `graphs` и т.д.). При формулировке запроса для ввода знака доллара может быть использована клавиша `Tab`.

В качестве запроса может быть введено поисковое выражение, полученное комбинированием двух или более поисковых терминов с г-

пользованием логических операторов, определяющих отношения между терминами. Допускается использование трех операторов: логическое OR (знак +), логическое AND (знак *) и логическое NOT (знак ^), а также круглых скобок.

Если задаваемый в запросе поисковый термин содержит круглые скобки или специальные символы ("*", "+", "~", ".", "\$", "#"), то необходимо выделить его знаками двойного апострофа (""), чтобы избежать неоднозначности и ошибок во вводимом запросе.

В процессе выполнения операции поиска система BLESC выдает на экран компьютера краткий оперативный протокол работы. Ниже на рисунке 1 приведен фрагмент такого протокола, полученного при обработке поискового запроса "адрон\$ * модел\$", а также расшифровка и описание основных его элементов.

```

Номер запроса: 3: АДРОН$ * МОДЕЛ$
P= 2 АДРОН (1)
P= 7 АДРОНОГО
P= 2 АДРОННОЙ
  o o o
P= 22 АДРОНОВ
P= 1 АДРОНЫ
T= 45 - #12: АДРОН$ (2)
P= 13 МОДЕЛЕЙ (3)
P= 120 МОДЕЛИ
P= 51 МОДЕЛИРОВАНИЕ
  o o o
P= 1 МОДЕЛЯМИ
P= 10 МОДЕЛЯХ
T= 262 - #30: МОДЕЛ$ (4)
T= 4 - #31: #12 * #30 (5)
T= 4 - #3: #31 (6)
  
```

Запрос: АДРОН\$ * МОДЕЛ\$ (Число записей: 4)

Номер запроса: 3

<p>Жданов Г.Б., Денисова В.Г. Анализ основных характеристик гамма-адронных семейств. Сравнение эксперимента с расчетными моделями для углеродных камер. #ФИАН, Москва, 1991, 114, 14с. ил. Библиогр.-с.14. [С 347 л # 118/93]</p> <p>Аккелин С.В., Мартинов Е.С. Кроссинг-нечетные спиновые эффекты в дипольной модели рассеяния адронов. #1992 - т.55, вып.11, ЯФ, с.3013-23. Библиогр.-17. [С 323.2]</p>
--

{Enter} - Следующая страница

{Esc} - Завершить просмотр

Рис. 1. Фрагмент протокола на экране компьютера при обработке поискового запроса.

Первая строка протокола выполнения операции поиска всегда содержит номер запроса, присваиваемый системой, и сам запрос, выраженный прописными символами. Строки, начиная с (1), содержат количества публикаций (P=pppp), имеющих в своих библиографических реквизитах термины, начинающиеся с символов "адрон". Строка (2) содержит итоговое количество записей (T=45) и присвоенный номер поискового подвыражения (#12) для выражения "адрон\$". Аналогично строки, начиная с (3), и строка (4) содержат соответствующие сведения (T=262 и #30) для выражения "модел\$". Строка (5) содержит количество библиографических записей (T=4) для пересечения (логическое AND) подвыражений #12 и #30, то есть для выражения "адрон\$ * модел\$". Последняя строка (6) содержит итоговое значение количества найденных записей (T=4), а также присвоенные системой номер поискового выражения запроса (#3) и номер подвыражения (#31) из строки (5). Естественно, что наибольший интерес для пользователя представляет число публикаций, удовлетворяющих его поисковому запросу. Если это число в последней строке протокола не равно нулю, то есть найдена хотя бы одна запись в базе данных, то после выдачи протокола инициируется просмотр библиографических ссылок отобранных публикаций.

5.1 Варианты простых запросов

- Найти публикации, в библиографических реквизитах которых присутствует термин "катализ":

Запрос: катализ

- Найти публикации, в библиографических реквизитах которых присутствует термин "quark":

Запрос: quark

- Найти публикации, автором которых является Петров М.И.:

Запрос: Петров М.И.

- Найти публикации на русском и английском языках, автором которых является Иванов А.Н.:

Запрос: иванов а.н. + ivanov a.n.

- Найти публикации, в библиографических реквизитах которых присутствуют термины, начинающиеся с символов "резонанс":

Запрос: резонанс\$

- Найти публикации, автором которых является Сидоров (точнее, ВСЕ Сидоровы – независимо от инициалов):

Запрос: Сидоров\$

- Найти публикации, в библиографических реквизитах которых присутствуют термины, начинающиеся с символов "адрон" И с символов "модел":

Запрос: адрон\$ * модел\$

- Найти препринт с регистрационным номером 4375, полученный НТВ в 1994 году:

Запрос: 4375/94

- Найти публикации по лептонам, изданные на русском и английском языках в Протвино:

Запрос: лептон\$ * протвино + lepton\$ * protvino

- Найти публикации по предметной рубрике "С 324.2" (нелокальные и нелинейные теории поля), изданные в 1994 году:

Запрос: "с 324.2" * 1994

- Найти публикации, в библиографических реквизитах которых присутствует термин LHC ИЛИ термины, начинающиеся с символов "trigger", НО отсутствуют термины, начинающиеся с символов "test":

Запрос: (LHC + trigger\$) ^ test\$

5.2 Варианты сложных запросов

- Поиск по близости терминов применяется для ограничения выборки публикаций посредством задания логически конкретизирующих условий в запросе, он особенно полезен для нахождения некоторых естественных фраз или словосочетаний. Так, если А и В – два поисковых термина, то использование в запросе точки, выделенной с обеих сторон пробелами, позволяет удовлетворить следующим требованиям при поиске:

А . В – термины расположены по соседству (рядом);

А . . В – не более одного термина между А и В;

А . . . В – не более двух терминов между А и В;

и т.д.

Найти публикации, в библиографических реквизитах которых присутствуют расположенные РЯДОМ термины, начинающиеся с символов "graph" И с символов "analys":

Запрос: graph\$. analys\$

- Уточнение поискового запроса заключается в использовании результатов предыдущих операций поиска при формулировании запросов к базе данных. Как было упомянуто выше в пояснениях к рисунку 1, система присваивает уникальный номер каждому элементу поискового запроса. Она также сохраняет библиографические записи, соответствующие каждому пронумерованному элементу запроса. Поэтому в последующих запросах можно ссылаться на предыдущие, используя номера самих запросов или их элементов, перед которыми необходимо указывать знак номера (#). Это средство позволяет формировать запрос итерационно, шаг за шагом сужая или расширяя множество отобранных публикаций. Разбиение запроса на элементы дает возможность не только проверять на каждом шаге количество найденных записей, но и выдавать на экран компьютера сами записи.

Из множества публикаций, найденных в 8-ом запросе, исключить записи, в библиографических реквизитах которых присутствуют слова, начинающиеся с символов "test":

Запрос: #8 ^ test\$

К публикациям, соответствующим поисковому подвыражению с номером 31, добавить записи, в библиографических реквизитах которых присутствуют слова "CERN" И "CMS":

Запрос: #31 + CERN * CMS

- Поиск по свободному тексту заключается в простом последовательном переборе указанных библиографических реквизитов записей на предмет выполнения заданных критериев. Этот вид поиска не использует аппарат инвертированных списков CDS/ISIS, в связи с чем является очень неэффективным и может быть рекомендован только в крайнем случае, когда исчерпаны все поисковые возможности, рассмотренные выше. Последовательный поиск обозначается символом вопросительного знака (?) в начале запроса и производится либо по всей базе данных, либо только по одной из предшествующих выборок, указанной ее номером (с символом #). Поисковый критерий задается при помощи идентификатора библиографического реквизита (v10 – поле авторов, v20 – поле заглавия публикации), символа двоеточия (:), означающего функцию включения, и искомого значения для сравнения. В процессе выполнения поиска по свободному тексту на экране компьютера высвечивается оперативный протокол со следующими параметрами:

MFN – номер текущей обрабатываемой записи;
Hits – количество отобранных записей;
% – процент отобранных записей к общему количеству;
Recs – общее число обрабатываемых записей.

В любой момент времени можно приостановить выполнение поиска по свободному тексту и посмотреть частичный результат.

Найти публикации, в поле авторов которых имеются фамилии, содержащие символы "idor":

Запрос: ? v10 : 'idor'

Среди публикаций, отобранных в запросе номер 7, найти записи, в поле заглавия которых присутствуют символы 'lider':

Запрос: ? #7 v20 : 'lider'

6. УСТАНОВКА СИСТЕМЫ BLESC

Для упрощения установки на компьютер пользователя интерфейса с информационно-поисковой системой BLESC созданы специальные процедуры (командные файлы), полностью автоматизирующие этот процесс. На подключаемом компьютере требуется наличие программного обеспечения для взаимодействия по протоколу IPX с сервером LIB НТБ ОИЯИ посредством вычислительной сети ETHERNET.

На предложение ввести имя пользователя при установлении связи с файл-сервером необходимо ответить 'blesc', при этом не требуется предъявлять пароль для вхождения в рабочий сеанс связи:

Enter your login name: blesc

Командные файлы, используемые для установки программного обеспечения клиента, имеют наименования соответственно inst_c.bat и inst_d.bat – в зависимости от выбора жесткого диска C или D в качестве рабочего. Указанные файлы находятся на файл-сервере в каталоге F:\PROG. Рабочий диск используется для размещения временных файлов в процессе выполнения поисковых запросов, он должен быть доступен для операций записи. Чтобы избежать проблем, связанных с нехваткой места на диске, желательно обеспечить на нем свободное пространство размером не менее 1 Мб.

В процессе установки на выбранном рабочем диске создается каталог \BLESC, куда помещаются все необходимые файлы, среди которых командная процедура blesc.bat, используемая в дальнейшем для запуска информационно-поисковой системы НТБ ОИЯИ. В этом же каталоге затем будут размещаться сохраняемые пользователем результаты обработки его поисковых запросов. Чтобы обеспечить вызов системы BLESC, независимо от текущего расположения, желательно скопировать файл blesc.bat в один из каталогов, включенных в системный поисковый путь PATH в системном файле autoexec.bat.

7. ЗАКЛЮЧЕНИЕ

При создании информационно-поисковой системы BLESC ставилась задача реализации максимально простого и интуитивно ясного диалога пользователя в сочетании с предоставлением эффективного и мощного языка запросов. Имеющийся опыт эксплуатации показывает, что з

основном эту задачу удалось успешно решить: любой пользователь после нескольких минут знакомства начинает уверенно работать с системой, а время ее реакции на запрос удовлетворит самого взыскательного пользователя: на обработку достаточно сложных поисковых запросов порою требуется менее 1 секунды. И это при том, что объем накопленной в НТБ ОИЯИ информации составляет уже около 70 Мегабайт.

Несомненной технологической удачей является успешная интеграция системы BLESS и автоматизированных рабочих мест сотрудников НТБ ОИЯИ, что обеспечило функционирование всего комплекса на единой информационной базе данных. Регулярное, обычно еженедельное, пополнение поисковой базы данных BLESS выполняется непосредственно сотрудниками библиотеки, в адрес которых особенно хочется сказать много теплых слов. Именно благодаря доброжелательности, профессионализму и самоотверженности этого коллектива во главе с искренне болеющей за свое дело Т.И.Харжесовой удается поддерживать на высоком уровне информационно-библиотечное обслуживание научных сотрудников ОИЯИ.

Безусловно, полученные результаты представляют собой лишь первые шаги в создании современного информационного обеспечения научных исследований. Вместе с тем следует признать, что разработанные средства удаленного коллективного доступа к библиографическим данным НТБ предоставляют информационный сервис качественно нового уровня. В настоящее время изучаются вопросы, связанные с хранением полных текстов публикаций, по крайней мере, поступающих из издательского отдела ОИЯИ, рассматриваются проблемы доступа к различным внешним библиографическим базам данных, а также возможности оперативного обмена сведениями о последних поступивших публикациях между научно-техническими библиотеками ОИЯИ и ЦЕРН.

ЛИТЕРАТУРА

1. Пакет прикладных программ CDS/ISIS/M. Версия 3.0. - М., Международный центр научной и технической информации, 1992.

Рукопись поступила в издательский отдел
12 октября 1994 года