

ОРГАНИЗАЦИЯ ИНФОРМАЦИОННОЙ РАБОТЫ

УДК 002:004.658 ВИНТИ

Т. М. Леонтьева, А. В. Рудикова, А. Н. Шогин

О новых услугах для пользователей Банка данных ВИНТИ и средствах их реализации

Рассматриваются отличия новой версии ИПС банка данных (БнД) ВИНТИ — Федерального банка данных по научно-технической информации.

Основные изменения, внесенные с целью повышения эффективности и комфортности работы с БнД ВИНТИ, состоят в новой схеме диалога пользователя с ИПС, новых функций, таких, как “История поиска” и “Архив запросов”, новых форматов вывода документов.

Предложенные решения основываются на современных методах построения поисковых страниц и выходных форм, использующих средства DHTML и CSS для динамического формирования фреймовых структур; на создании новой версии системы контроля работы пользователей, предусматривающей использование протокола с установлением соединения без квитирования; на применении средств обмена, базирующихся на новейших стандартах описания гипертекстовых документов и протоколов обмена данными.

С целью повышения эффективности и комфортности работы с банком данных (БнД) ВИНТИ предлагаются новые решения по организации пользовательского интерфейса и дополнительные функции ИПС, в том числе:

- новая схема диалога пользователя с поисковой системой;
- новая функция “История поиска” как средство временного хранения запросов для отладки и составления сложных запросов;
- новая функция “Архив запросов” как средство долговременного хранения запросов;
- новые формы вывода найденных документов с возможностью изменения состава и способа представления выводимых данных.

Старая схема диалога [1–4] предусматривала две страницы для ввода запросов — “Поиск” и “Расширенный поиск”, а также отдельную страницу для вывода сведений о количестве найденных документов вместе с первой порцией самих документов. Страница “Поиск” была предназначена для составления простых запросов, а страница “Расширенный поиск” — для построения сложных запросов. Как показала практика, такое разбиение не учитывало всех вариантов потребностей пользователей.

При разработке новой схемы диалога было решено:

- объединить две поисковые страницы в одну страницу “Поиск” с полным набором инструментов для ввода ограничений поиска;
- ввести на страницу “Поиск” сведения о количестве найденных документов;
- объединить страницы “Словарь” и результаты поиска по словарю;

- предоставить пользователям возможность составления сложных запросов из подзапросов, используя вновь созданные инструменты “История поиска” и “Архив запросов”.

Новый вариант пользовательского интерфейса предоставляет возможность на одной странице “Поиск” ввести запрос и получить сведения о количестве найденных документов, исключая формирование страницы с самими документами (рис. 1). Это позволяет пользователю сразу оценить результат поиска и принять решение — редактировать запрос и вновь проводить поиск или посмотреть найденные документы, для чего каждый запрос снабжен гиперссылкой “показать”, позволяющей вывести на экран найденные документы. Для помощи пользователям при составлении или редактировании запросов на страницу “Поиск” введены списки дополнительных поисковых признаков, реализованные в виде всплывающих окон. Используя элементы списка “Поле” (рис. 2), можно ограничить поиск по запросу или просмотр словаря только заданными элементами данных. Использование элементов списка “Выпуски” (пример на рис. 3) позволяет ограничить тематическую область поиска, включив в запрос имя и значение (шифр) тематического раздела (выпуска) выбранной базы данных (БД).

Из списка “Год, №” (рис. 4) можно включить в запрос конкретный год (из пятилетней ретроспективы) и месяц генерации БД.

Из списка “Язык, страна...” (см. рис. 4) запрос может быть дополнен кодом страны, языком и годом издания, а также видом первоисточника.

Подобное изменение схемы диалога реализовано и для страницы “Словарь”, использование которой

помогает отбирать слова или их фрагменты при подготовке запросов, предоставляя информацию о встречаемости (количестве документов и частоте) этих слов в БД. В новой версии на одной странице можно ввести искомое слово (или фрагмент) и получить список найденных словоформ (см. рис. 2). На этой же странице в отдельном окне можно посмотреть документы, содержащие эту словоформу, поскольку каждая из них является гиперссылкой. Кроме того, расширены возможности поиска по словарю. Теперь можно ограничить просмотр не одним, а сразу несколькими полями.

Новая функция "История поиска" в основном предназначена для отладки запросов и формирования сложных запросов, но может использоваться и как средство хранения текущих двадцати запросов. Можно сначала провести поиск по нескольким простым запросам, а затем, оценив релевантность, скомбинировать их, используя возможности "Истории поиска" (см. рис. 1). В "Истории поиска" представлена следующая информация о выполненном запросе:

- текст запроса;
- дата проведения поиска;

Запросы	Дата поиска	БД	Найдено
<input type="checkbox"/> 20((ЭНЕРГОСБЕРЕЖ\$ AND ЭЛЕКТРОЭНЕРГ\$) IN AB, TI, KW) AND 2004\$ IN ID AND (AHLJ) IN LA)	2004-11-29	EN2001-2004	3 показать
<input type="checkbox"/> 19((ЭНЕРГОСБЕРЕЖ\$ AND ЭЛЕКТРОЭНЕРГ\$) IN AB, TI, KW) AND 2004\$ IN ID AND (RU) IN CC)	2004-11-29	EN2001-2004	146 показать
<input checked="" type="checkbox"/> 18(ЭНЕРГОСБЕРЕЖ\$ AND ЭЛЕКТРОЭНЕРГ\$)	2004-11-29	EN2001-2004	984 показать
<input type="checkbox"/> 17((ЭНЕРГОСБЕРЕЖ\$ AND ЭЛЕКТРОЭНЕРГ\$) IN AB, TI, KW)	2004-11-29	EN2001-2004	405 показать
<input type="checkbox"/> 16(((ЭНЕРГОСБЕРЕЖ\$)) IN AB, TI, KW)	2004-11-29	EN2001-2004	4674 показать
<input type="checkbox"/> 15(ЭНЕРГОСБЕРЕЖ\$)	2004-11-29	EN2001-2004	6251 показать
<input type="checkbox"/> 14(((ЭНЕРГОСБЕРЕЖ\$)) IN SO)	2004-11-29	EN2001-2004	2621 показать
<input type="checkbox"/> 13(ЭНЕРГОСБЕРЕЖ\$ IN SO)	2004-11-29	EN2001-2004	2621 показать
<input type="checkbox"/> 12(((ЭНЕРГОСБЕРЕЖ\$ OR ЭНЕРГОСБЕРЕЖЖ\$) AND ЭЛЕКТРОЭНЕРГ\$) IN AB, TI, KW) AND 2004\$ IN ID AND (RU) IN CC)	2004-11-29	EN2001-2004	146 показать

Рис. 1

Словарь [Просмотр](#) [Удалить запрос](#)

База данных: [Выбор БД](#) ?

Энергетика 2001-2004

Слово: [Поле](#) ?

Искать в: [Поле](#) ?

Список найденных словоформ, содержащих фрагмент: "ЭНЕРГОСБЕРЕЖ"

ЭНЕРГОСБЕРЕЖАТЕЛЬ (1,1)
ЭНЕРГОСБЕРЕЖАТЕЛЬЯМИ (1,1)
ЭНЕРГОСБЕРЕЖАЮЩАЯ (101,109)
ЭНЕРГОСБЕРЕЖАЮЩЕГО (106,127)
ЭНЕРГОСБЕРЕЖАЮЩЕЕ (46,50)
ЭНЕРГОСБЕРЕЖАЮЩЕЙ (128,148)
ЭНЕРГОСБЕРЕЖАЮЩЕМ (3,3)
ЭНЕРГОСБЕРЕЖАЮЩЕМУ (1,1)
ЭНЕРГОСБЕРЕЖАЮЩИЕ (626,709)
ЭНЕРГОСБЕРЕЖАЮЩИЙ (67,71)
ЭНЕРГОСБЕРЕЖАЮЩИМ (27,27)
ЭНЕРГОСБЕРЕЖАЮЩИМИ (18,18)
ЭНЕРГОСБЕРЕЖАЮЩИХ (659,768)
ЭНЕРГОСБЕРЕЖАЮЩЮЮ (6,6)
ЭНЕРГОСБЕРЕЖ (7,7)
ЭНЕРГОСБЕРЕЖ-НА-ВАШЕМ-ПРЕДПРИЯ (5,5)
ЭНЕРГОСБЕРЕЖЕНИЕ (4246,5502)
ЭНЕРГОСБЕРЕЖЕНИЕ+КОМФОРТ-КВАРТ (1,1)
ЭНЕРГОСБЕРЕЖЕНИЕ-1999 (1,1)
ЭНЕРГОСБЕРЕЖЕНИЕ-2000 (4,6)
ЭНЕРГОСБЕРЕЖЕНИЕ-2001 (8,17)
ЭНЕРГОСБЕРЕЖЕНИЕ-2002 (2,2)
ЭНЕРГОСБЕРЕЖЕНИЕ-2003 (2,3)
ЭНЕРГОСБЕРЕЖЕНИЕ-В-ПОВОЛЖЬЕ (250,250)
ЭНЕРГОСБЕРЕЖЕНИЕ-В-САРАТОВ-ОБЛ (170,170)
ЭНЕРГОСБЕРЕЖЕНИЕ-И-ВОДОПОДГОТ (320,320)
ЭНЕРГОСБЕРЕЖЕНИЕ-И-ПРОБЛ-ЭНЕРГ (71,71)
ЭНЕРГОСБЕРЕЖЕНИЕ/АКК (1,1)
ЭНЕРГОСБЕРЕЖЕНИЕ] (2,2)
ЭНЕРГОСБЕРЕЖЕНИЕМ (145,170)
ЭНЕРГОСБЕРЕЖЕНИИ (132,145)
ЭНЕРГОСБЕРЕЖЕНИЙ (4,4)
ЭНЕРГОСБЕРЕЖЕНИЮ (436,513)
ЭНЕРГОСБЕРЕЖЕНИЮ] (1,1)

Рис. 2

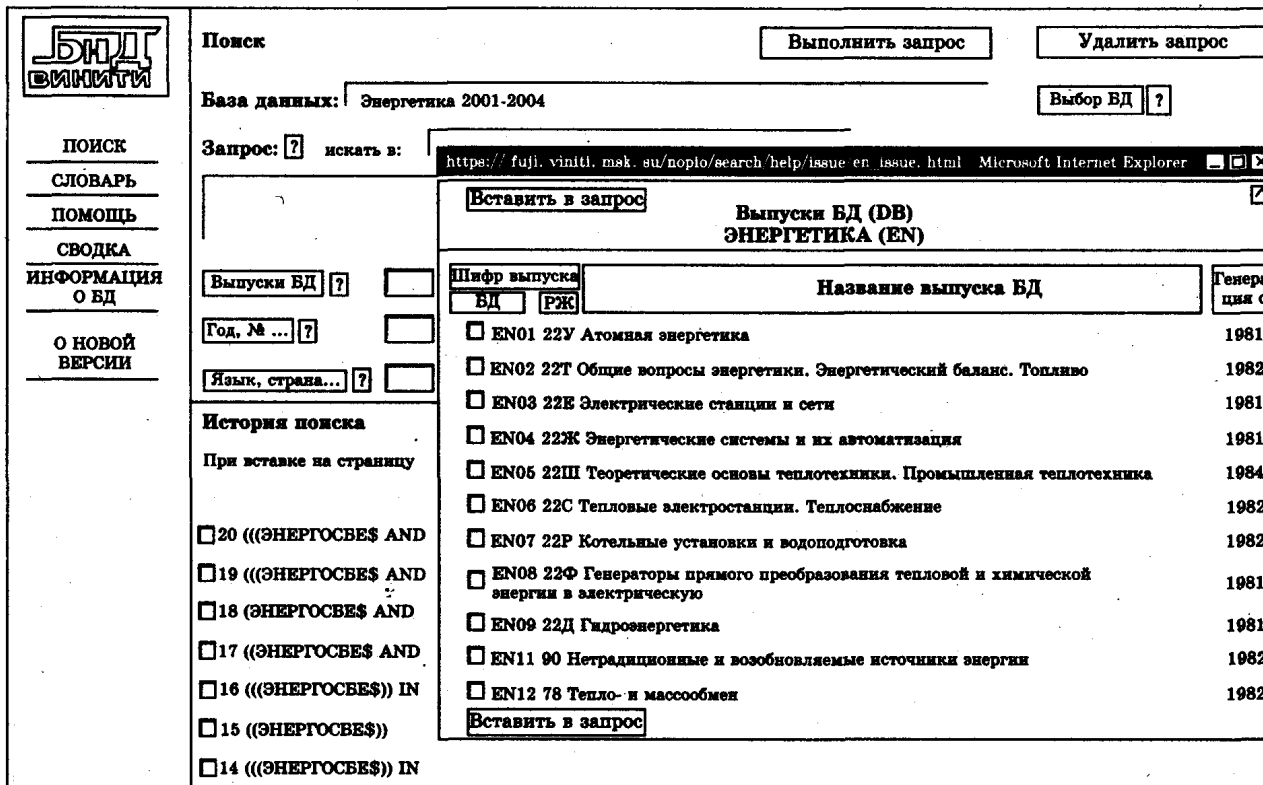


Рис. 3

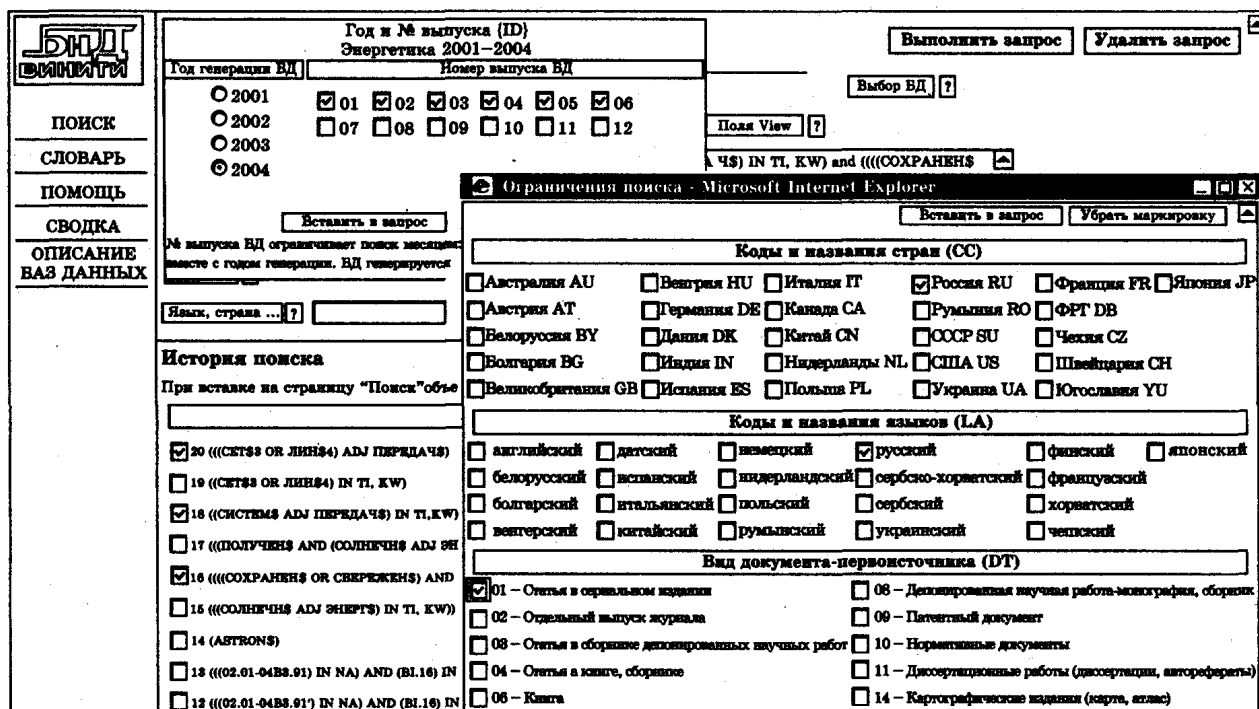


Рис. 4

- префикс базы данных, в которой проводился поиск;

- количество найденных документов.

Каждый запрос имеет свой чекбокс, при помощи которого пользователь маркирует необходимые для комбинирования запросы, выбирает нужный оператор ("И", "ИЛИ") и затем вставляет выбранные запросы в основное окно "Запрос" страницы "Поиск", образуя новый сложный запрос.

Запросы в "Историю поиска" записываются автоматически (как только проведен поиск) и замещаются автоматически (двадцать первый запрос вытесняет бывший первый и так далее).

Для обеспечения возможности длительного хранения запросов, что необходимо, в первую очередь, для пользователей, регулярно работающих с постоянными запросами, создан механизм архивизации запросов.

Страница "Архив запросов" (рис. 5) похожа по структуре на страницу "История поиска", но дополнена полями "Имя запроса" и "Примечание". Эти поля, заполняемые пользователем самостоятельно, не являются обязательными. Однако их наличие полезно для дальнейшего использования и организации архива.

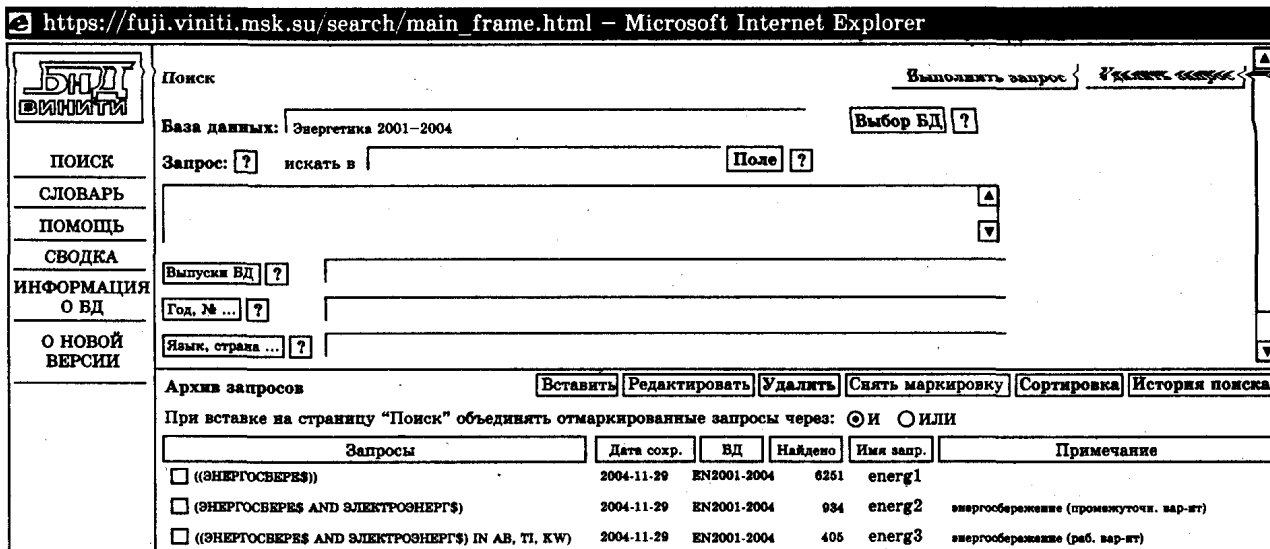


Рис. 5

Чтобы поместить запрос в архив, пользователь маркирует его в “Истории поиска” и запускает диалог сохранения в архиве.

“Архив запросов” позволяет переносить хранимые запросы (один или несколько) в поле “Запрос” на странице “Поиск” (механизм переноса аналогичен используемому в “Истории поиска”), удалять отмаркированные запросы или все сразу, редактировать дополнительно введенные данные и упорядочивать по этим данным список запросов.

Кроме вышеописанных основных изменений, характеризующих новую версию ИПС БНД ВИНТИ, был внесен в ряд других, обеспечивающих повышение эффективности и комфортности работы пользователей. Остановимся на одном из них — совершенствовании форм вывода документов.

По желанию пользователей добавлена форма “Библиография”, которая может использоваться как готовый материал при подготовке списков литературы.

Введена новая форма вывода, в которой состав элементов данных (полей), включаемых в описание документов, и способ их представления, задаются пользователем — форма “Заданные поля” (рис. 6).

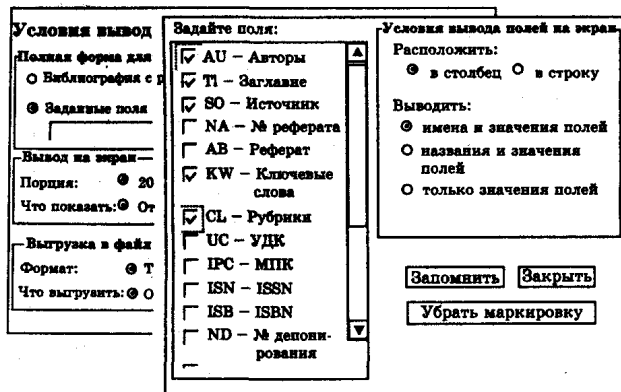


Рис. 6

Пользователь может выбрать любой набор полей из предлагаемого списка и, кроме того, задать способ их расположения (в столбец, в строку), а также вариант отображения выбранных полей — с названием (префикс или полное название поля) или без названия (только значение поля).

Настройки условий вывода документов, заданные пользователем, запоминаются системой. Таким образом, один раз задав условия вывода, пользователь может просматривать результаты поиска в удобном для него виде.

Большинство вышеописанных изменений было реализовано посредством решения задачи персонализации пользовательских данных (запросов и некоторых других параметров) в системе биллинга. Изначально в ней не было предусмотрено хранение детальной информации о действиях пользователя. Для решения этой задачи в биллинговую систему был введен ряд таблиц для персональной информации: user_query — хранение запросов, user_quotas — хранение квот пользователей и другие. Эти таблицы не являются OLAP-таблицами и предназначены для долговременного хранения информации.

Потребовалась также разработка следующих программно-технологических средств на стороне сервера:

- создание вторичного сервера биллинговой системы для SQL доступа к специализированным таблицам биллинга;
- модификация системы обработки шаблонов для приема — передачи SQL запросов к биллинговой системе на стороне клиента;
- модификация API поисковой системы для передачи SQL запросов к биллинговой системе.

Кроме того, в биллинговую систему был введен механизм квотирования доступа пользователей. Последнее важно для корпоративных пользователей с ограничением доступа.

Рассмотрим все эти компоненты подробно.

Вторичный сервер биллинговой системы запускается первичным сервером непосредственно после генерации OLAP таблицы. В процессе запуска производится обмен информацией между первичным и вторичным серверами для синхронизации открытия ими портов доступа. Такая синхронизация обеспечивает корректное функционирование всех клиентов биллинговой системы. Непосредственно после старта вторичный сервер биллинга начинает принимать SQL запросы (через определенный в конфигурации порт) в формате SQL <текст запроса>.

Поскольку во все внутренние таблицы биллинга введены схемы разграничения доступа, доступ пользователя возможен только к своим данным. Доступ же пользователя, не зарегистрированного в базе данных, блокирован полностью из-за использования шифрования трафика по алгоритму SSI. Дополнительная аутентификация не производится, поскольку доступ обслуживается лишь с тех хостов, которые заданы в конфигурации.

Реализация приема результатов выполнения запросов потребовала значительной модификации протокола взаимодействия "клиент — сервер" из-за резкого увеличения и усложнения ответа сервера. Протокол передачи запроса может быть либо UDP с квитированием, т. е. с подтверждением передачи дейтаграммы, либо TCP — протокол с установлением соединения без квитирования. Последний используется для передачи запросов, отличных от выборок.

Ответ сервера зависит от типа запроса — все запросы, кроме запроса SELECT, сопровождаются квитанцией об успешном или неуспешном завершении транзакции. В случае запроса на выборку в ответ посылаются ряд квитанций об успешно выбранных данных. Последней является простая квитанция об окончании выборки. В любой момент может поступить квитанция об аварийном завершении выборки. Отдельные элементы выборки разделяются символом ^ . Данный символ не используется в базах данных и в текстах запросов и поэтому легко может быть использован для разделения данных.

Вторичный сервер биллинга сам по себе завершить свою работу не может. Для этого ему необходимо принять сообщение QUIT от первичного сервера. Это сделано для синхронизации остановки серверов биллинга. Более того, периодически, первичный и вторичный серверы связываются друг с другом и, если связи не происходит, завершают свою работу.

Следует также отметить, что непосредственное взаимодействие с базой данных биллинга осуществляется на уровне API. Таким образом, все протокольные связи являются полностью платформо-независимыми.

Были выполнены также и разработки в клиентской части, касающиеся формирования SQL запросов и приема (при необходимости) ответов вторичного сервера биллинга. Во фронтальной части системы поиска взаимодействие с биллингом осуществляется на двух уровнях. Низкий уровень обеспечивает непосредственную передачу и прием сообщений, высокий — трансляцию данных, введенных (или неявно формируемых) пользователем, в SQL запросы и обратную трансляцию SQL выборок в вид, пригодный для вывода на экран.

Низкий уровень интегрирован в API поисковой системы и доступен из любых программ, использующих это API. На этом уровне разработаны программные блоки приема и передачи сообщений вторичному серверу биллинга в определенных ситуациях. Набор этих ситуаций описывается следующими условиями: а) принудительная передача любого SQL сообщения, б) обязательная передача сообщения о записи результатов поиска в "Историю поиска", в) передача сообщения о записи по запросу пользователя в нестандартных ситуациях.

Фактически в API введена одна функция — она обеспечивает формирование внутреннего протокольного кода SQL, передачу его вторичному

серверу биллинга, прием ответа и разбор его с одновременной генерацией внутреннего массива данных, который используется далее механизмом шаблонов для вывода информации на экран пользователя.

На высоком уровне модифицирован механизм использования шаблонов. Можно сказать, что в настоящее время этот механизм обеспечивает полнофункциональный доступ к базам данных биллинга. Он состоит из ряда функций: передача произвольного SQL запроса, контроль результатов его исполнения, прием и разбор результатов выборки. С помощью этого механизма доступ к биллинговой информации легко обеспечивается через шаблоны вывода данных и, тем самым, через выводимые HTML страницы.

К новым функциям механизма шаблонов относятся:

- передача произвольного SQL запроса вторичному сервису биллинга;
- вывод указанной позиции (строка, колонка) результатов обработки запроса или пусто, если позиция отсутствует;
- выдача количества строк в выборке;
- освобождение динамической памяти, выделенной под выборку.

В принципе, использование последней функции не является обязательным, поскольку после окончания короткоживущего процесса обработки шаблона эта память все равно будет освобождена. Однако при обработке насыщенных SQL запросами страниц, ее все же лучше использовать. Приведем фрагмент шаблона для вывода запросов пользователя:

```
*mysql_query('select qno,query,qdb,substing(qtime,1,10),qrecs from user_query where userid='\'+getenv('REMOTE_USER')+'\'and qno<=20 order by qno desc')*
.....
${for(1;mysql_cnt;1)}
.....
mark_edit(1);DB_name='*mysql_res(_i0,3)*';DltSum('*mysql_res(_i0,1)*');
if(this.checked==false){
mark_edit(2);}">

```

Отметим, что несмотря на название функций, они никак не связаны с СУБД MySQL и могут с равным успехом быть использованы для любой СУБД. Это связано с автономностью протокола взаимодействия между вторичным сервером биллинга и клиентами.

В настоящее время новая версия ИПС внедрена в эксплуатацию.

СПИСОК ЛИТЕРАТУРЫ

1. Егоров В. С., Леонтьева Т. М., Шогин А. Н. База данных ВИНТИ и ее использование в сети INTERNET и на CD // Материалы 6-й Международной конференции НТИ.— М.: ВИНТИ, 2002.— С. 129-130.