

РЕАЛИЗАЦИЯ АЛГОРИТМА ОБРАБОТКИ ДАННЫХ ДЛЯ ПРЕДОСТАВЛЕНИЯ ЭФФЕКТИВНОГО ИНФОРМАЦИОННОГО ПОИСКА В БАЗЕ ДАННЫХ ПУБЛИКАЦИЙ

Бахмад Э.А., Курочкина Е.В., Королева И.Ю.

ВолГТУ «Волгоградский государственный технический университет», (400005, Россия, г. Волгоград, пр. Ленина 28), bahmad.ilya@gmail.com

Многие библиотеки высших учебных заведений ведут базу данных публикаций сотрудников, которая поддерживается в локальной сети библиотеки. Поэтому была поставлена задача: реализовать интернет-версию раздела «Картотека публикаций». В данном разделе должны быть сформированы различные виды фильтров поиска и отчетов по разным типам публикаций (15 видов публикаций, разбитых на 7 групп). Так как база данных была разработана под систему MS DOS и хранит данные о публикациях в формате *.dbf, было необходимо разработать алгоритм конвертирования из формата *.dbf в формат MS SQL. При реализации алгоритма конвертирования удалось в несколько раз сократить время генерации результирующей таблицы. Конечная информация хранится в таблице, работа с которой во многом упрощена, что способствует повышению эффективности и скорости обработки поисковых запросов. Также необходимо было разработать алгоритм для формирования библиографических списков, получаемых в результате поиска, в соответствии с действующим стандартом.

Ключевые слова: электронная библиотека, библиографическое описание, публикации сотрудников, научная техническая библиотека, информационный поиск в базах данных, электронная библиотечная система.

THE IMPLEMENTATION OF THE DATA PROCESSING ALGORITHM TO PROVIDE EFFECTIVE INFORMATIONAL SEARCH IN DATABASE OF PUBLICATIONS

Bahmad I.A., Kurochkina E.V., Koroleva I.Y.

VSTU «Volgograd State Technical University» (400005, Russia, Volgograd, Lenin avenue 28), bahmad.ilya@gmail.com

Many libraries of higher educational institutions work with publications database of their employees in the local network. Thus I was given the task to implement the Internet version of the section “Card Index of Publication”. In this section, different search filters and reports on various types of publications (15 types of publications, divided into 7 groups) will be carried out. Since the database with which I worked, was developed under DOS and the data of publications in the *.dbf format, it was necessary to develop an algorithm to convert from the *.dbf format to MS SQL format. The implement of the algorithm for converting, managed to reduce the time generation of the resulting table by several times. The final data is stored in a resulting table. Working with this resulting table is simplified thereby increasing efficiency and speed query processing. It was also necessary to develop an algorithm to generate bibliographic lists in accordance with the applicable standard derived from the search.

Key words: digital library, bibliographic description, publications of staff, scientific and technical library, informational search in databases, e-library system.

Введение

База данных публикаций сотрудников высших учебных заведений представляет собой уникальную коллекцию библиографических записей. Она успешно используется университетом с целью содействия развитию научных школ и стимулирования труда профессорско-преподавательского состава, сотрудников и аспирантов.

В результате анализа процесса работы пользователей базы данных публикаций было принято решение разработать интернет-версию раздела «Картотека публикаций». В данном

разделе должны быть реализованы различные виды фильтров поиска и отчетов по разным типам публикаций. Такая версия работы раздела является востребованной для всех категорий читателей и наиболее эффективна при использовании.

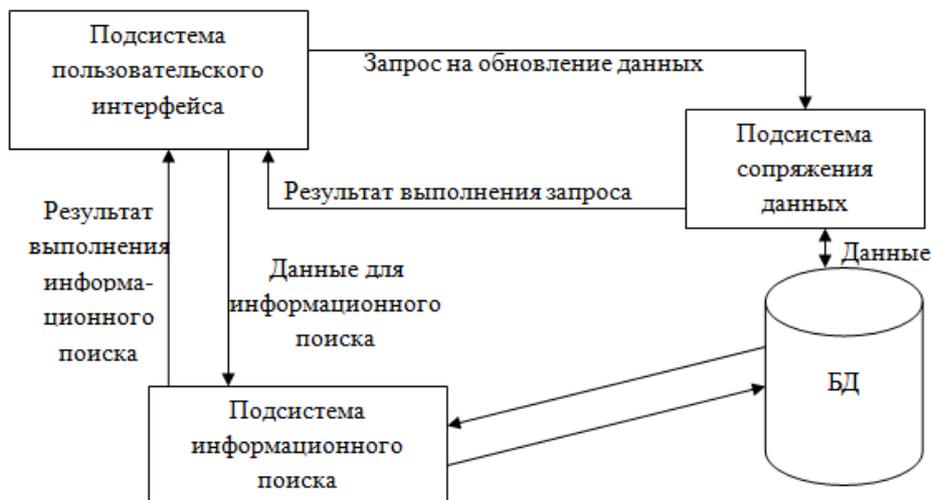


Рис. 1. Архитектура библиотеки.

Информационный обмен между компонентами системы осуществляется посредством программно-информационной сетевой среды. Подсистемы взаимодействуют между собой посредством вызова функций (рис. 1).

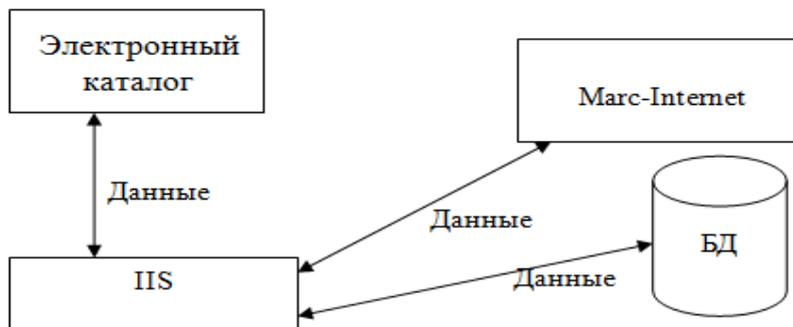


Рис. 2. Взаимосвязь электронного каталога со смежными системами.

Электронный каталог взаимодействует со смежными подсистемами через Internet Information Server (IIS), используя платформу Framework.NET (рис. 2) [5].

Загрузка базы данных публикаций

Изначально база данных публикаций разрабатывалась под систему MS DOS. Данные в ней хранятся в формате *.dbf в виде набора из 13 таблиц.

В связи с форматом исходной базы данных, а также большим числом записей о публикациях (около 60000) было необходимо разработать наиболее эффективный способ преобразования данных.

В результате работы был сформирован алгоритм, осуществляющий соединение с базой данных и перенос информации из DBF-таблиц в соответствующие таблицы MS SQL. Были

произведены внутренние преобразования базы данных MS SQL [1]. Часть данных была получена с помощью вызовов динамически формируемых SQL-запросов. Другая часть – с помощью программной обработки в памяти, с использованием компонента DataSet для получения итоговой таблицы «_Katalog» [2].

Таблица «_Katalog» хранит сведения об источнике, достаточные для выполнения информационного поиска. Она связана с таблицей «library» полем «reg_n». В таблице «library» хранятся библиографические карточки (рис. 3).

Для преобразования информации понадобилось большое количество промежуточных таблиц, из которых была сформирована результирующая таблица. Для формирования результирующей таблицы было необходимо осуществить несколько операций объединения. Все операции объединения организованы в виде SQL-запросов, что значительно ускоряет обработку данных.

В целом алгоритм загрузки базы данных публикаций выглядит следующим образом.

1. Загрузка файлов DBF на сервер.
 - 1.1. Подмена бита кодировки.
2. Подключение к базе данных на DBF-файлах.
3. Перенос данных из DBF в MS SQL.
4. Преобразования над данными для получения результирующей таблицы «_Katalog».
 - 4.1. Очистка вспомогательных таблиц «KATALOG_N» и «FAK_KAF».
 - 4.2. Выборка упорядоченных по индексу публикаций из таблицы «KATALOG».
 - 4.3. Выборка упорядоченных по индексу авторов из таблицы «FIO_SP».
 - 4.4. Проход по всем публикациям (все записи выборки из таблицы «KATALOG»)
 - 4.4.1. Формирование списка авторов, кафедр и факультетов.
 - 4.4.1.1. Добавление первого автора из таблицы «KATALOG» и соответствующих автору аббревиатуры кафедры и факультета в списки из таблиц «FAK» и «KAFEDRA».
 - 4.4.1.2. Проход по таблице соавторов «FIO_SP», предварительно занесённой в «DataSet» и дополнение списков аббревиатурами кафедр и факультетов из таблиц «FAK» и «KAFEDRA».
 - 4.4.2. Формирование таблицы «KATALOG_N» из списков авторов.
 - 4.4.3. Формирование таблицы «_FAK_KAF» из списков кафедр и факультетов.
 - 4.5. Формирование таблицы «KATALOG_N2» (SQL-запросом), путём объединения таблиц «KATALOG» и сформированной таблицы «KATALOG_N».
 - 4.6. Формирование результирующей таблицы «_KATALOG» (SQL-запросом), путём объединения таблиц «KATALOG_N2» и сформированной таблицы «_FAK_KAF».

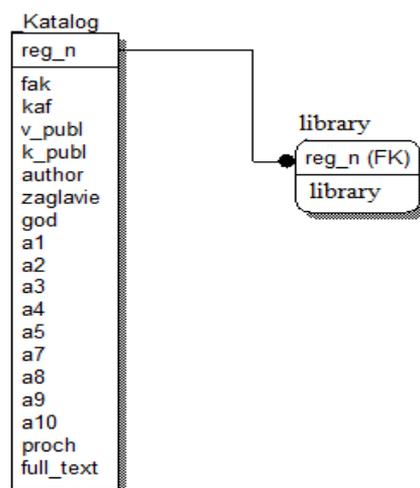


Рис. 3. Структура БД публикаций.

В результате реализации данного алгоритма удалось в несколько раз сократить время генерации результирующей таблицы, а также снизить временные затраты сервера.

Алгоритм формирования библиографического списка

При выполнении поиска публикаций в базе данных необходимо сначала сформировать его библиографическое описание. Оно создается по ГОСТ 7.1 – 2003 и хранится в отдельной таблице. Такой ход необходим для увеличения производительности системы при поиске [1]. Если результатом поиска будет более 100 документов, и каждый из них требует библиографического описания, то при создании библиографических карточек «на лету» время, затрачиваемое на выполнение одного запроса, существенно возрастает.

Создание библиографического описания (БО) полнотекстового документа выполняется по следующему алгоритму [4].

1. Если текущий документ является монографией, учебником или учебным пособием, то
 - 1.1. Если число авторов ≤ 3 , то
 - 1.1.1. Если «ФИО авторов» $\neq \langle \rangle$, то
 БО = «ФИО авторов» + «.»
 - 1.1.2. Если «Название книги» $\neq \langle \rangle$, то
 БО = БО + «Название книги» + «:»
 - 1.1.3. Если «Сведения, поясняющие название (учебник, пособие, методические материалы)» $\neq \langle \rangle$, то
 БО = БО + «Сведения, поясняющие название (учебник, пособие, методические материалы)» + «/»
 - 1.2. Если число авторов > 3 , то
 - 1.2.1. Если «Название книги» $\neq \langle \rangle$, то

БО = «Название книги» + «:»

1.2.2. Если «Сведения, поясняющие название (учебник, пособие, методические материалы)» != «», то

БО = БО + «Сведения, поясняющие название (учебник, пособие, методические материалы)» + «/»

1.2.3. Если «ФИО авторов» != «», то

БО = БО + «ФИО авторов» + «;»

1.3. Если «ФИО редактора, переводчика, название организации, ответственной за подготовку книги» != «», то

БО = БО + «ФИО редактора, переводчика, название организации, ответственной за подготовку книги» + «.»

1.4. Если «Повторность издания (2-е издание, переработанное и дополненное)» != «», то

БО = БО + «-» + «Повторность издания (2-е издание, переработанное и дополненное)»

1.5. Если «Место издания» != «», то

БО = БО + «-» + «Место издания» + «:»

1.6. Если «Название издательства» != «», то

БО = БО + «Название издательства» + «,»

1.7. Если «Год издания» != «», то

БО = БО + «Год издания» + «.»

1.8. Если «Общее количество страниц книги» != «», то

БО = БО + «-» + «Общее количество страниц книги» + «.»

2. Если текущий документ является статьей из российского или зарубежного журнала, то

2.1. Если число авторов <=3, то

2.1.1. Если «ФИО авторов» != «», то

БО = «ФИО авторов» + «.»

2.1.2. Если «Название статьи» != «», то

БО = БО + «Название статьи» + «//»

2.2. Если число авторов >3, то

2.2.1. Если «Название статьи» != «», то

БО = «Название статьи» + «/»

2.2.2. Если «ФИО авторов» != «», то

БО = БО + «ФИО авторов» + «//»

2.3. Если «Название журнала» != «», то

БО = БО + «Название журнала» + «.»

- 2.4. Если «Год издания» != «», то
 БО = БО + «-» + «Год издания» + «.»
- 2.5. Если «Номер журнала, год или номер выпуска» != «», то
 БО = БО + «-» + «Номер журнала, год или номер выпуска» + «.»
- 2.6. Если «Страницы, на которых опубликована статья» != «», то
 БО = БО + «-» + «Страницы, на которых опубликована статья» + «.»
3. Если документ является статьей из сборника, тезисом доклада или известиями, то
- 3.1. Если число авторов <=3, то
- 3.1.1. Если «ФИО авторов» != «», то
 БО = «ФИО авторов» + «.»
- 3.1.2. Если «Название статьи» != «», то
 БО = БО + «Название статьи» + «//»
- 3.2. Если число авторов >3, то
- 3.2.1. Если «Название статьи» != «», то
 БО = «Название статьи» + «//»
- 3.2.2. Если «ФИО авторов» != «», то
 БО = БО + «ФИО авторов» + «//»
- 3.3. Если «Название сборника» != «», то
 БО = БО + «Название сборника» + «.»
- 3.4. Если «Место издания» != «», то
 БО = БО + «-» + «Место издания» + «.»
- 3.5. Если «Год издания» != «», то
 БО = БО + «Год издания» + «.»
- 3.6. Если «Страницы, на которых опубликована статья» != «», то
 БО = БО + «-» + «Страницы, на которых опубликована статья» + «.»

Алгоритм разбора строки запроса для поиска в базе данных публикаций

Для выполнения поиска в базе данных публикаций необходимо было разработать алгоритм разбора строки записи. Строка записи имеет вид: "name = Название документа <> author = Автор <> PlaceIzd = Место издания <> Year = Год издания <> Str = Количество страниц".

Для разбора строки была использована функция Split [3].

Функция Split("title", "LogOP", A)

Title – Поле, соответствующее определенной части запроса;

LogOP – логический оператор;

A – элемент массива.

Пока текущий символ строки «a» не равен “”, делать

Если текущий символ строки «a» не равен “;”, то

tmp=title(i)

Иначе

res = res & Title & " like '%" & tmp & "%"

res = res & " " & LogOP & " "

Возвращаем преобразованную строку

Split = res

SQL-запрос формируется так.

1. Разбираем строку записи:

1.1. Создаем массив «a» из 5 элементов

1.2. Пока количество символов в строке запроса не равно 0, делать

1.2.1. Если текущее поле запроса == “name”, то

1.2.1.1. a(1) = Название документа

1.2.1.2. a(1) = Split("BookValue.FieldValue", "AND", a(1))

1.2.2. Если текущее поле запроса == “author ”, то

1.2.2.1. a(2) = Автор

1.2.2.2. a(2) = Split("BookValue.FieldValue", "AND", a(2))

1.2.3. Если текущее поле запроса == “PlaceIzd ”, то

1.2.3.1. a(3) = Место издания

1.2.3.2. a(3) = Split("BookValue.FieldValue", "AND", a(3))

1.2.4. Если текущее поле запроса == “ Year”, то

1.2.4.1. a(4) = Год издания

1.2.4.2. a(4) = Split("BookValue.FieldValue", "AND", a(4))

1.2.5. Если текущее поле запроса == “ Str”, то

1.2.5.1. a(5) = Количество страниц

1.2.5.2. a(5) = Split("BookValue.FieldValue", "AND", a(5))

2. Составляем строку запроса:

2.1. Пока число элементов в массиве «a» не равно 5, делать

2.1.1. Если текущий элемент массива «a» не равен “”, то

2.1.1.1. Если строка запроса не равна “”, то

2.1.1.1.1. selSQL = selSQL & " OR "

2.1.1.1.2. Составляем строку запроса selSQL = selSQL & " (" & a(i)
& ") "

Заключение

Практическая ценность работы заключается в предоставлении информационного поиска в базе данных публикаций посредством Интернета и в максимальном повышении эффективности этого поиска.

Список литературы

1. Бахмад Э.А. Фундамент электронных библиотек / Э.А. Бахмад, И.Ю. Королева // Журнал научных публикаций аспирантов и докторантов. – 2011. – № 11. – С. 92-95.
2. Белых А.В. Нормализация объектно-ориентированных баз данных на основе N-модели данных / А.В. Белых, С.М. Ковалев, О.В. Ольховик // Известия ВолгГТУ. Серия «Актуальные проблемы управления вычислительной техники и информатики в технических системах» : межвуз. сб. науч. ст. / ВолгГТУ. – Волгоград, 2009. – Вып. 7. – № 12. – С. 58-61.
3. Королева И.Ю. Проектирование баз данных : учеб. пособие. – Волгоград : ВолгГТУ, 2008. – 31 с.
4. Методические указания по применению ГОСТ 7.1-2003 «Библиографическая запись. Библиографическое описание. Общие требования и правила составления» / сост. Н.Н. Аржановская; ред. И.М. Рамзина, Н.С. Заруднева. – Волгоград : ВолгГТУ, 2009. – 18 с.
5. Тейбор К. Реализация XML Web-служб на платформе Microsoft NET : пер. с англ. – М. : Вильямс, 2002. – 448 с.

Рецензенты:

Лукьянов В.С., д.т.н., профессор, зав. кафедрой «ЭВМ и системы», Волгоградский государственный технический университет, г. Волгоград.

Камаев В.А., д.т.н., профессор, зав. кафедрой «САПР и ПК», Волгоградский государственный технический университет, г. Волгоград.