

На правах рукописи

ШИВАРОВ АЛЕКСАНДР ЕВГЕНЬЕВИЧ

**СОЗДАНИЕ СИСТЕМЫ ОРГАНИЗАЦИИ ДОСТУПА К
ИНФОРМАЦИОННЫМ РЕСУРСАМ
УНИВЕРСИТЕТСКОЙ БИБЛИОТЕКИ**

**Специальность 05.13.11 – Математическое и программное
обеспечение вычислительных машин, комплексов и
компьютерных сетей**

АВТОРЕФЕРАТ

**диссертации на соискание ученой степени
кандидата технических наук**

Москва – 2007

Работа выполнена в Московском государственном техническом университете им. Н.Э. Баумана.

Научный руководитель: доктор технических наук, профессор
Трусов Борис Георгиевич

Официальные оппоненты: доктор технических наук, профессор
Антопольский Александр Борисович
кандидат технических наук, доцент
Филиппович Юрий Николаевич

Ведущая организация: Государственная публичная научно-
техническая библиотека России

Защита диссертации состоится «__» _____ 200_ года на заседании диссертационного совета Д 212.141.10 в Московском государственном техническом университете им. Н.Э. Баумана по адресу: 107005, Москва, 2-я Бауманская ул., д. 5.

С диссертацией можно ознакомиться в библиотеке Московского государственного технического университета им. Н.Э. Баумана.

Автореферат разослан «__» _____ 200_ г.

Ученый секретарь
диссертационного совета Д 212.141.10

к.т.н., доцент

С.Р. Иванов

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы

Одной из первостепенных задач, стоящих перед университетской библиотекой сегодня, является совершенствование методов и средств доступа читателей к информационным ресурсам. Начавшаяся во второй половине прошлого века автоматизация библиотек и создание автоматизированных библиотечно-информационных систем (АБИС) сделали возможным использование электронных каталогов взамен традиционных карточных, повышая скорость и удобство поиска информационных ресурсов. Тем не менее, во многих существующих АБИС электронный каталог является лишь автоматизированной версией традиционного каталога, позволяющей быстрее находить информацию по запросам за счет хранения информационного массива в памяти вычислительной машины.

Анализ эффективности использования современных электронных каталогов АБИС показал, что существующие подходы к организации пользовательского интерфейса предоставляют читателям ограниченные возможности доступа к информационным ресурсам. Основной причиной низкой эффективности является преимущественное использование поискового метода организации доступа пользователя к каталогу. При использовании поискового метода организации доступа взаимодействие пользователя с системой происходит в основном по сценарию «запрос-ответ», заставляя пользователя постоянно переформулировать поисковый запрос, что требует определенного опыта использования информационно-поисковых систем и знаний поискового языка конкретной системы. Перечисленные обстоятельства являются причиной низкой эффективности функционирования электронных каталогов для пользователей с недостаточным опытом, что особенно актуально для читателей университетской библиотеки.

Альтернативой поисковому методу организации доступа является поисково-навигационный подход, который позволяет пользователю во многих случаях избежать сложной процедуры формулирования поискового запроса за счет системы навигации между связанными разделами каталога, рубриками и отдельными ресурсами. Использование подобных механизмов позволяет сделать информационно-поисковую систему для библиотечных ресурсов доступной для пользователей с различным уровнем подготовки в области информационного поиска. Таким образом, разработка поисково-навигационного метода организации доступа к информационным ресурсам библиотеки представляется актуальной проблемой.

Целью работы является разработка поисково-навигационного метода организации доступа к информационным ресурсам университетской библиотеки и программная реализация данного метода в рамках подсистемы

автоматизированной библиотечно-информационной системы библиотеки университета.

Для достижения этой цели в диссертации решены следующие **задачи**:

- Предложен и реализован поисково-навигационный метод организации доступа к информационным ресурсам библиотеки университета;
- Разработаны методы построения ассоциативных связей между разделами каталога информационных ресурсов и отдельными его записями для обеспечения функционирования средств навигации;
- Создана подсистема АБИС, реализующая поисково-навигационный метод организации доступа к информационным ресурсам.

Методы исследования.

В диссертации для разработки поисково-навигационного метода организации доступа к информационным ресурсам применяются методы теории вероятностей, математической статистики, теоретико-множественный аппарат.

При создании подсистемы АБИС используются методы декомпозиции системы на уровни и слои, методы моделирования информационных систем при помощи универсального языка моделирования, методы объектно-ориентированного программирования с использованием сред разработки Microsoft Visual Studio.NET и СУБД Microsoft SQL Server.

Научная новизна.

В результате выполненной работы получены новые научные результаты:

- предложен новый метод организации доступа к информационным ресурсам библиотеки университета на основе поисково-навигационного подхода;
- разработаны методы построения ассоциативных связей между разделами каталога информационных ресурсов и отдельными его записями для усовершенствования в нем системы навигации;
- предложены критерии и процедура оценки, позволяющие определять эффективность поисково-навигационного метода организации доступа к массиву информационных ресурсов.

Основные положения, выносимые на защиту:

- поисково-навигационный метод организации доступа к информационным ресурсам библиотеки университета;
- метод построения ассоциативных связей между разделами каталога информационных ресурсов и отдельными его записями для усовершенствования в нем системы навигации;
- критерии и процедура оценки эффективности поисково-навигационного метода организации доступа к массиву информационных ресурсов;
- архитектура и алгоритмы функционирования подсистемы АБИС,

реализующей разработанный поисково-навигационный метод организации доступа к информационным ресурсам библиотеки университета.

Практическая значимость.

Предложенный в диссертации подход позволяет повысить качество обслуживания пользователей библиотечной системы за счет использования средств навигации для организации доступа к информационным ресурсам.

Основным практическим результатом работы является создание подсистемы АБИС, реализующей предложенный подход.

Апробация и внедрение результатов работы.

Созданная в результате выполненной работы подсистема организации доступа к информационным ресурсам внедрена и функционирует в рамках АБИС библиотеки МГТУ им. Н.Э. Баумана «Яуза».

Основные результаты работы докладывались и обсуждались на конференциях «Технологии Microsoft в теории и практике программирования 2005, 2006», десятой и одиннадцатой международных конференциях «LIBCOM-2006» и «LIBCOM-2007», научно-практическом семинаре «Новые информационные технологии в автоматизированных системах» (МИЭМ, 2007) и четырнадцатой международной библиотечной конференции «Крым 2007».

Публикации по теме диссертации.

Основные результаты диссертационной работы опубликованы в девяти печатных трудах. В том числе одна работа опубликована в издании, включенном в перечень ведущих рецензируемых научных изданий ВАК. Получены два свидетельства об официальной регистрации программ для ЭВМ.

Структура и объем диссертации. Диссертация состоит из введения, четырех глав, заключения и библиографического списка. Диссертация изложена на 133 страницах текста, содержит 36 рисунков и 6 таблиц. Библиография состоит из 97 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** изложена проблема организации доступа к информационным ресурсам университетской библиотеки, обоснована ее актуальность, сформулирована цель исследования и необходимость разработки поисково-навигационного метода организации доступа к массиву информационных ресурсов и создания подсистемы АБИС, реализующей данный метод.

В **первой главе** проведен анализ особенностей функционирования библиотеки университета и ее АБИС, детально рассмотрены вопросы

организации подсистем электронных каталогов, проведен обзор литературных источников по вопросам создания информационно-поисковых систем и их оценки.

В качестве особенностей функционирования библиотеки университета, определяющих требования к системе организации доступа к информационным ресурсам, выделены

- 1) первоочередная ориентация на обслуживание пользователей;
- 2) интеграция с учебным процессом;
- 3) множественность и гетерогенность источников ресурсов;
- 4) наличие значительного контингента малоопытных пользователей.

Одной из ключевых задач функционирования АБИС является подготовка и представление пользователям электронного каталога информационных ресурсов. Первые электронные каталоги появились в 60-70-х годах прошлого века и были «побочным продуктом» совершенствования библиотечных технологий. Фактически системы первого поколения представляли собой автоматизированные версии карточных каталогов, предоставляя пользователям те же возможности, что и карточные каталоги, но только с использованием ЭВМ. Для второго поколения электронных каталогов характерно применение методов булевого поиска терминов по полям библиографической записи. Это расширило возможности систем второго поколения по сравнению с первым. Однако эффективность существующих систем доступа к ресурсам библиотеки продолжает оставаться низкой, в особенности при использовании их малоопытными пользователями.

Проведенный анализ пользовательских интерфейсов современных АБИС российских разработчиков («ИРБИС», ГПНТБ России и Ассоциация ЭБНИТ, «Библиотека» и «Эйдос», Библиотечная Компьютерная Сеть, «МАРК», Информ-система, «Руслан» и «Нева», Открытые библиотечные системы, «ОРАС», ДИТ-М) показал, что указанные недостатки организации доступа к информационным ресурсам в той или иной степени имеют место в существующих разработках.

Рассмотрены классические принципы организации информационно-поисковых систем: теоретико-множественные и векторные модели информационного поиска, процедуры предварительной обработки текста. Проанализированы известные методы оценки информационно-поисковых систем, включая показатели полноты, точности, графики полноты-точности, макрооценки показателей точности и полноты.

Сделаны выводы о необходимости разработки нового метода организации доступа к информационным ресурсам библиотеки на основе поисково-навигационного подхода, а также реализации данного метода в виде соответствующей подсистемы АБИС библиотеки университета.

Во **второй главе** излагается поисково-навигационный метод организации доступа к информационным ресурсам, а также методы построения ассоциативных связей между разделами каталога и отдельными его записями.

Выделены два вида взаимодействия пользователя с системой доступа к информационным ресурсам: поисковое и поисково-навигационное, различающиеся способом передачи информации о потребности пользователя системе. В случае поискового вида взаимодействия пользователь формулирует запрос к системе на ее поисковом языке. При навигационном взаимодействии пользователь выбирает один из предложенных системой вариантов перехода в новое состояние (например, к просмотру содержимого некоторого раздела или подробного описания некоторой записи). Помимо поискового и навигационного возможно комбинированное поисково-навигационное взаимодействие, объединяющее их возможности путем чередования актов обоих видов взаимодействия. Именно такой подход положен в основу предлагаемого метода организации доступа к информационным ресурсам.

Для формализованного описания поисковой системы достаточно определить функцию релевантности $Rl(q,d)$, позволяющую вычислять количественную степень релевантности документа d запросу q . Для системы, реализующей поисково-навигационный метод организации доступа, определения функции релевантности недостаточно. По этой причине был предложен специальный подход к описанию подобных систем.

Введены понятия именованного подмножества записей массива информационных ресурсов как упорядоченной пары $\langle t, D \rangle$ (t – имя подмножества, D – подмножество записей каталога) и состояния взаимодействия пользователя с системой как упорядоченной тройки $\langle q, E, B \rangle$ (q – поисковый запрос на языке запросов системы, определяющий текущий набор представляемых пользователю записей, E – множество релевантных записей, отбираемых по запросу q и отображаемых пользователю, B – множество именованных релевантных подмножеств, отображаемых пользователю в качестве дальнейших вариантов уточнения запроса).

Система организации доступа к информационным ресурсам формально определена в виде упорядоченной пары $\langle S, P \rangle$ (S – множество возможных состояний взаимодействия пользователя с системой, P – функция перехода в пространстве состояний взаимодействия). Функция перехода P позволяет определить новое состояние взаимодействия пользователя в зависимости от текущего состояния и выполненного действия

$$s_2 = P(s_1, a), \quad (1)$$

где s_1 – исходное состояние взаимодействия, s_2 – состояние взаимодействия после действия пользователя, a – действие пользователя.

Действием пользователя a может быть выбор одной из релевантных записей, именованного подмножества или ввод поискового запроса непосредственно:

$$a \in E \cup B \cup Q, \quad (2)$$

где a – действие в некотором состоянии, E – множество отображаемых

записей для состояния s , B – множество отображаемых именованных подмножеств для состояния s , Q – множество поисковых запросов, которые может ввести пользователь.

На основе анализа состава информационных ресурсов библиотеки университета, а также характера потребностей ее пользователей, были предложены следующие виды именованных подмножеств:

- 1) подмножества записей, составляющих содержание определенной предметной рубрики, рубрики ГРНТИ или индекса УДК;
- 2) подмножества записей, имеющих общие значения некоторых полей библиографического описания (автор, издательство, серия);
- 3) подмножества записей, включенных в списки рекомендаций по отдельным учебным курсам;
- 4) подмножества ассоциативно связанных записей.

Для построения подмножеств записей первых трех видов необходимо произвести выборки записей из массива информационных ресурсов по определенным признакам, являющимися общими для каждого из видов.

Помимо подмножеств, образованных общими значениями отдельных атрибутов, существенную роль для системы навигации играют ассоциативные связи записей. Для каждой записи массива ресурсов могут быть найдены записи, ассоциативно связанные с ней по тому или иному косвенному признаку (частая совместная выдача, просмотр, совместное присутствие в результатах поиска). Для обнаружения таких связей записей разработаны процедуры обработки имеющихся данных об использовании ресурсов в системе. Построение подмножеств записей всех типов и ассоциативных связей, а также их обновление по мере наполнения фондов библиотеки обеспечивает подсистема доступа к информационным ресурсам.

Для определения количественной меры ассоциативной связи введено понятие степени ассоциативной связи записи d_1 с записью d_2 , равной вероятности p события, состоящего в том, что ресурс d_2 удовлетворяет некоторую информационную потребность при условии, что ее удовлетворяет d_1 .

$$a(d_1, d_2) = p(F(d_2) | F(d_1)) = P_{2,1}, \quad (3)$$

где $a(d_1, d_2)$ – степень ассоциативной связи d_1 с d_2 , $F(d_1)$, $F(d_2)$ – события, связанные с удовлетворением некоторой информационной потребности ресурсами d_1 и d_2 .

Для обеспечения возможности практического использования ассоциативных связей записей было проведено эмпирическое оценивание вероятности $P_{2,1}$, что потребовало выделения источников реальных событий $F(d)$ (признаков ассоциирования записей). Учитывая специфику библиотеки университета, были предложены следующие виды признаков:

- заказы читателей;
- данные о выдаче печатных документов читателям;
- данные о просмотре пользователями подробной информации о записях

при работе с системой;

- данные о присутствии записей в списках результатов выполнения поисковых запросов пользователей.

Эмпирическая оценка вероятности (3) рассмотрена на примере признака совместной выдачи. Один сеанс обслуживания s представляет собой множество документов, выданных читателю при единовременном посещении библиотеки. Все множество сеансов выдачи (произведенных в системе за все время эксплуатации) обозначено S , а множество сеансов, содержащих некоторый документ $d - S(d)$.

Применение методов математической статистики позволило найти доверительный интервал параметра $P_{2,1}$ биномиального распределения с вероятностью $100(1-\alpha)$. При выполнении условий (4) становится возможна замена эмпирического распределения случайной величины точечной оценки (5) нормальным распределением и определение границ доверительного интервала по формулам (6). В случае, если условия (4) не выполняются, доверительный интервал вычисляется по формулам (7) для случая $0 < m_w < m$, (8) для $m_w = 0$ и (9) для $m_w = m$.

$$\begin{cases} |S(d_1)|P_{2,1} > 5 \\ |S(d_1)|(1-P_{2,1}) > 5 \end{cases} \quad (4)$$

$$\tilde{P}_{2,1} = |S(d_1) \cap S(d_2)| / |S(d_1)| \quad (5)$$

$$P_{2,1}^{\pm} = \tilde{P}_{2,1} \pm z_{1-\frac{\alpha}{2}} \sqrt{\tilde{P}_{2,1}(1-\tilde{P}_{2,1})S(d_1)}, \quad (6)$$

где

$z_{1-\frac{\alpha}{2}}$ - $1-\frac{\alpha}{2}$ -я процентиль распределения $N(0,1)$.

$$\begin{aligned} P_{2,1}^- &= \frac{m_w}{m_w + (m - m_w + 1)F_{v_1^-, v_2^-}}, \\ P_{2,1}^+ &= \frac{(m_w + 1)F_{v_1^+, v_2^+}}{m - m_w + (m_w + 1)F_{v_1^+, v_2^+}}, \\ m &= |S(d_1)|, \\ m_w &= |S(d_1) \cap S(d_2)|, \\ v_1^- &= 2(m - m_w + 1), \\ v_2^- &= 2m_w, \\ v_1^+ &= 2(m_w + 1), \\ v_2^+ &= 2(m - m_w), \end{aligned} \quad (7)$$

где

$F_{a,b} - 1 - \frac{\alpha}{2}$ -я процентиль распределения Фишера со степенями свободы a и b .

$$P_{2,1}^- = 0, P_{2,1}^+ = 1 - \left(\frac{\alpha}{2}\right)^{\frac{1}{m}} \quad (8)$$

$$P_{2,1}^- = \left(\frac{\alpha}{2}\right)^{\frac{1}{m}}, P_{2,1}^+ = 1 \quad (9)$$

Помимо вычисления оценки значения условной вероятности проведено определение факта наличия зависимости между событиями $F(d_1)$ и $F(d_2)$ с использованием метода проверки статистических гипотез.

Так же, как для отдельных ресурсов, ассоциативные связи устанавливаются и для подмножеств. Особенно эффективными являются связи между подмножествами, образованными предметными рубриками или классификационными индексами, т.к. они позволяют пользователю переходить к просмотру содержимого похожих рубрик того же или другого рубрикатора. Ассоциативные связи именованных подмножеств могут быть заданы (например, существуют справочники соответствия индексов УДК и рубрик ГРНТИ) или построены автоматически системой.

Далее рассмотрена функция перехода P , отдельные компоненты которой обозначены P_q, P_E, P_B , отдельные компоненты упорядоченной тройки состояния взаимодействия s обозначены соответственно s_q, s_E, s_B .

$$s_{q2} = \begin{cases} a, a \in Q \\ Q(a), a \notin Q, \end{cases} \quad (10)$$

где s_{q2} – значение компонента q упорядоченной тройки $P(s_1)$, $Q(a)$ – функция построения запроса по записи или подмножеству (порождает запрос на поиск конкретной записи или всех записей подмножества), a – действие пользователя, Q – множество поисковых запросов.

Множество релевантных записей (компонента E) образуется путем выполнения поиска записей в соответствии с поисковым запросом нового состояния

$$s_{E2} = F(C, s_{q2}), \quad (11)$$

где s_{E2} – значение компонента E упорядоченной тройки $P(s_1)$, $F(C, q)$ – функция поиска записей в массиве C по запросу q .

Компонента s_B нового состояния определяется по следующей формуле

$$s_{B2} = \begin{cases} B^1(a), a \in E \\ B^*(s_{q2}), a \in Q \cup B, \end{cases} \quad (12)$$

где s_{B2} – значение компоненты B упорядоченной тройки $P(s_1)$, $B^1(c)$ – функция определения релевантных подмножеств для одной записи, $B^*(q)$ – функция определения подмножеств, релевантных заданному поисковому запросу.

$$B^1(c) = B_R(c) \cup B_A(c), \quad (13)$$

где

$B_R(c)$ – подмножества каталога, образованные значениями атрибутов записи c ;

$B_A(c)$ – подмножества записей, ассоциативно связанных с c .

В множество $B_R(c)$ входят подмножества записей, образованные значениями следующих атрибутов записи c :

- предметными рубриками,
- элементами индекса УДК,
- рубриками ГРНТИ,
- ответственными лицами (авторами, редакторами),
- издательством,
- серией,
- рекомендуемыми учебными курсами.

Процесс вычисления подмножеств, релевантных поисковому запросу (функция $B^*(q)$), зависит от наличия в нем тех или иных поисковых условий. В случае наличия условия поиска по ключевым словам в качестве релевантных подмножеств используются релевантные данному условию классификационные рубрики, определяемые при помощи предложенного в работе метода. В том случае, если поисковый запрос содержит условие поиска по заданной предметной рубрике, индексу УДК, рубрике ГРНТИ или какого-либо другого рубрикатора, в число релевантных подмножеств включаются ее подрубрики, ее надрубрики, рубрики, связанные с ней отношением «см. также», ассоциативно связанные с ней рубрики.

Разработанный поисково-навигационный подход в совокупности с методами построения ассоциативных связей позволяет спроектировать и реализовать подсистему доступа к информационным ресурсам университетской библиотеки, обладающую большей эффективностью, чем существующие системы электронных каталогов.

В **третьей главе** описана созданная в рамках диссертации подсистема доступа к информационным ресурсам АБИС «Яуза», включая описание архитектуры подсистемы и ее места в АБИС, структур данных и алгоритмов отдельных методов и процедур.

С точки зрения программной архитектуры АБИС представляет собой многоуровневую систему, где количество и состав уровней может быть различным для отдельных подсистем. Общая схема архитектуры АБИС библиотеки университета показана на рис. 1.

Подсистема доступа к информационным ресурсам обеспечивает представление ресурсов читателям и состоит из двух модулей: модуля представления ресурсов и модуля построения поискового образа. Общая схема

работы подсистемы показана на рис. 2. На рис. 3 показана архитектура модуля представления информационных ресурсов.

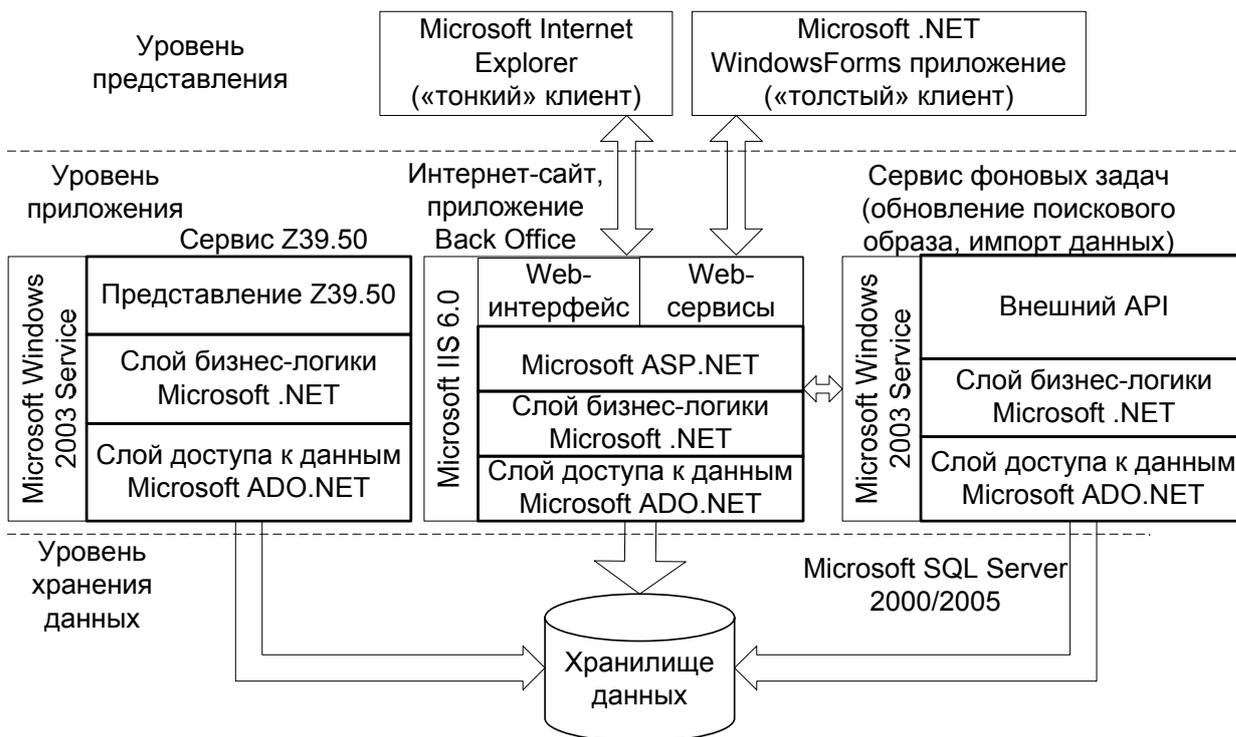


Рис. 1 Общая схема архитектуры АБИС университета

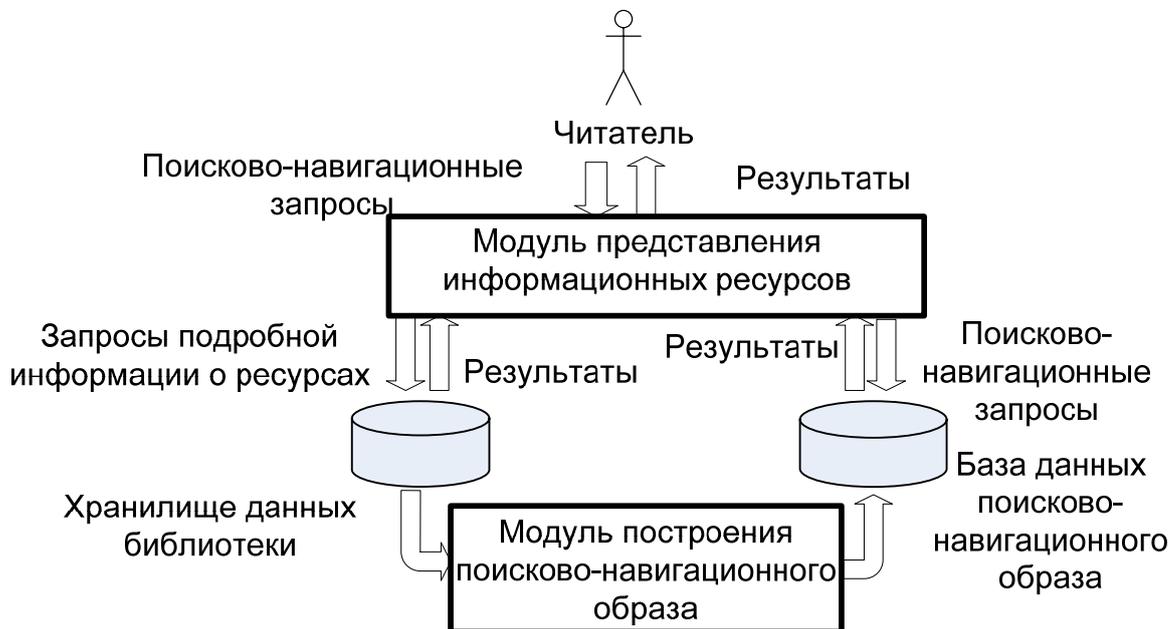


Рис. 2 Схема работы подсистемы доступа к информационным ресурсам библиотеки

Хранилище данных библиотеки наполняется в процессе ее комплектования печатными и электронными документами. Модуль построения поисково-навигационного образа приводит входящую информацию, получаемую в виде унифицированных описаний документов, к виду, удобному для эффективной работы алгоритмов поиска и навигации. Модуль представления электронных ресурсов осуществляет взаимодействие с пользователем (читателем): принимает запросы, выполняет поиск, представляет результаты в удобном виде, а также предоставляет возможности для навигации по электронному каталогу.

Поисково-навигационный образ представляет собой специальную структуру данных, позволяющую производить выборку необходимой информации для работы пользовательского интерфейса доступа к ресурсам. Для организации хранилища поисково-навигационного образа используется реляционная база данных.

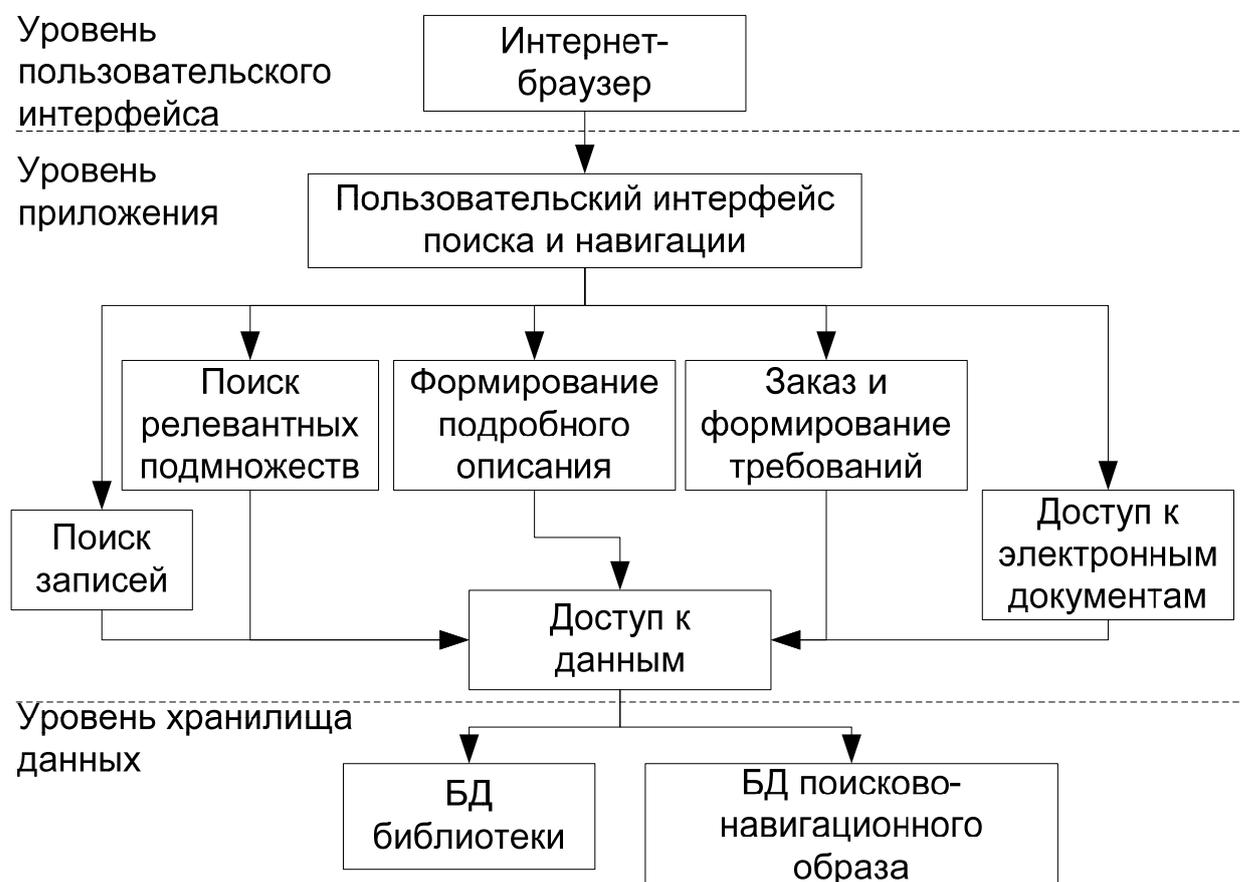


Рис. 3 Архитектура модуля поиска и представления информационных ресурсов

Созданная подсистема позволяет предоставлять доступ к информационным ресурсам библиотеки университета в комбинированном поисково-навигационном режиме, а также может быть использована для оценки

эффективности данного подхода по сравнению с традиционным.

В четвертой главе представлены разработанные процедуры и критерии оценки эффективности поисково-навигационного метода организации доступа к информационным ресурсам библиотеки университета, а также излагаются условия проводившихся экспериментов и их результаты.

Для оценки классических информационно-поисковых систем используется методика, при которой вычисляются показатели полноты и точности (или их производные) на специально подготовленных массивах поисковых данных. Для оценки поисково-навигационной системы доступа к информационным ресурсам разработано обобщение классического подхода.

В качестве основного массива экспериментальных данных используется база данных библиографических записей библиотеки МГТУ им. Баумана по состоянию на июнь 2007 года объемом около 100 000 записей.

Для оценки эффективности поисково-навигационного метода организации доступа в его реализации для АБИС МГТУ им. Н.Э. Баумана предложен ряд частных показателей эффективности, позволяющих оценить отдельные аспекты поисково-навигационного механизма, и обобщенный показатель эффективности метода в целом. Частные показатели позволяют оценить эффективность построения ассоциативных связей между записями и подмножествами записей, использования ассоциативных связей подмножеств с отдельными записями, поиска релевантных подмножеств. Обобщенный показатель позволяет оценить повышение эффективности поискового процесса в целом.

Для оценки значений предложенных показателей используется тестовый набор информационных потребностей, подготовленных экспертом для оценки эффективности поисково-навигационного метода организации доступа к информационным ресурсам. Каждая информационная потребность I в данном наборе представляет упорядоченную пару $I = \langle q, R \rangle$, где q – поисковый запрос, которым пользователь может выражать данную потребность, R – множество релевантных данной потребности записей.

Для оценки эффективности построения ассоциативных связей между записями рассчитывается доля $q(I, r)$ релевантных записей среди ассоциативно связанных с записью r в рамках некоторой информационной потребности I (14)

$$q(I, r) = \frac{1}{|R|} \sum_{r \in R} \frac{|A(r) \cap R|}{|A(r)|}, \quad (14)$$

где $A(r)$ – множество ассоциативно связанных записей для записи r .

Среднее значение показателя (14) для тестового набора составило 0,48, т.е. в среднем около половины записей, ассоциативно связанных с заданной (релевантной некоторой потребности), релевантны той же информационной потребности.

Для того чтобы оценить эффективность построенных ассоциативных связей между подмножествами записей, предложен показатель среднего относительного

увеличения количества релевантных записей $g(I,s)$ при просмотре содержимого подмножества s для заданной информационной потребности I (15)

$$g(I,s) = \frac{\left| \left(s \cup \bigcup_{t \in S_A} t \right) \cap R \right|}{|s \cap R|}, \quad (15)$$

где S_A – множество подмножеств, ассоциативно связанных с s .

Для тестового набора данных среднее значение (15) составило 2,46, т.е. добавление к подмножеству ассоциативно связанных с ним подмножеств увеличивает в среднем количество релевантных записей (для некоторой информационной потребности) в 2,5 раза.

Эффективность использования подмножеств, связанных с записью, оценивается путем расчета показателя полноты удовлетворения потребности I за счет навигации по подмножествам, связанным с записью r (16).

$$h(I,r) = \frac{\left| \bigcup_{t \in S_r} t \cap R \right|}{|R|}, \quad (16)$$

где S_r – множество подмножеств, связанных с записью r .

Среднее значение показателя (16) для тестового набора составило 0,58, т.е. связанные с записью, релевантной некоторой потребности, подмножества содержат в среднем около 60% релевантных той же потребности записей.

Для оценки эффективности поиска релевантных подмножеств предложен показатель относительного увеличения полноты удовлетворения потребности I для объединения релевантных запросу q подмножеств (17)

$$w(I,B) = \frac{\left| \left(E \cup \bigcup_{t \in B} t \right) \cap R \right|}{|E \cap R|}, \quad (17)$$

где B – множество подмножеств, релевантных q , E – множество записей, релевантных q .

Для тестового набора данных среднее значение (17) составило 1,45, т.е. добавление к результатам поиска по ключевым словам записей из релевантных тому же поисковому запросу подмножеств увеличивает полноту результатов в среднем на 45%.

Обобщенная оценка эффективности поисково-навигационной модели выполняется на основе графа навигации $G = \langle V, E \rangle$. Граф навигации строится для заданного поискового запроса q , его вершинами являются состояния взаимодействия пользователя с системой, достижимые из начального состояния,

соответствующего поиску по запросу q . Дугами графа навигации являются навигационные переходы (состоящие в выборе записи или подмножества) между состояниями взаимодействия (рис. 4). На графе результатов поиска вводится функция достижимости документов $\rho(d)$, равная длине минимального пути из начальной вершины в вершину документа d . Также рассматривается функция области достижимости $\delta(r)$ равная множеству документов, находящихся от начальной вершины на расстоянии менее или равном r .

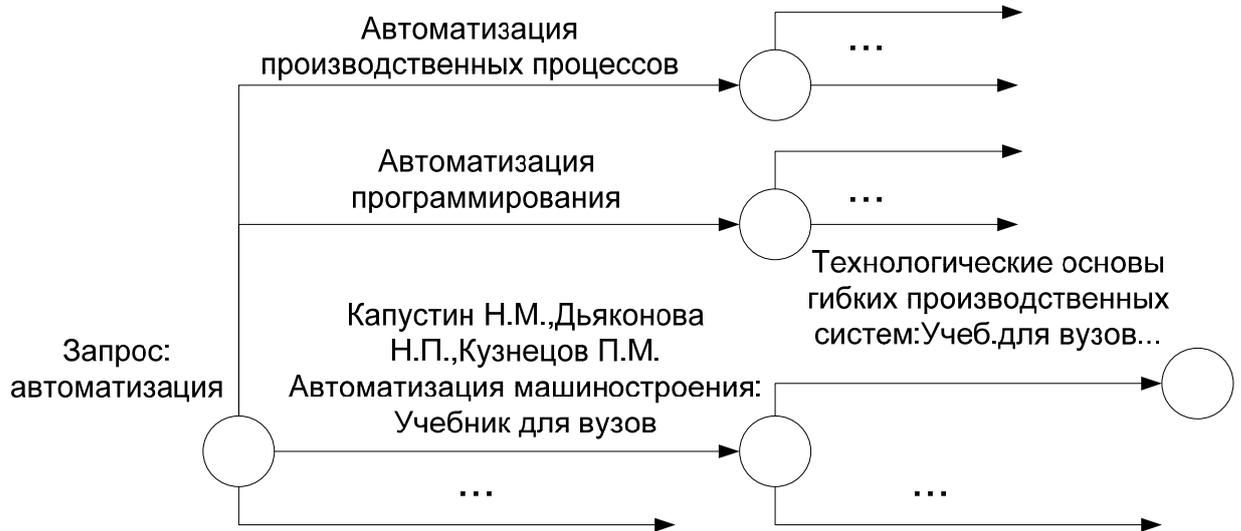


Рис. 4 Пример фрагмента графа навигации

Обобщенным показателем эффективности поисково-навигационной системы является показатель потенциальной полноты, определяемый для заданной информационной потребности I по формуле (18)

$$F(I, k) = \frac{|\delta(k) \cap R|}{|R|}, \quad (18)$$

где $F(I, k)$ – показатель потенциальной полноты в зависимости от информационной потребности и расстояния в графе результатов поиска.

Показатель потенциальной полноты показывает долю документов среди релевантных информационной потребности, которые могут быть найдены, в зависимости от расстояния в графе результатов поиска (в зависимости от количества переходов в пространстве состояний взаимодействия, что на практике означает число переходов по ссылкам в веб-интерфейсе электронного каталога).

Результаты оценивания показателя потенциальной полноты показаны на рис. 5. Гистограмма показывает сравнение значений показателя полноты для поискового и поисково-навигационного вариантов организации доступа к ресурсам. В случае поискового варианта роль расстояния играет количество просматриваемых страниц результатов поиска (при 10 записях на странице).

Проведенные на тестовом массиве данных эксперименты подтверждают эффективность предложенного подхода.

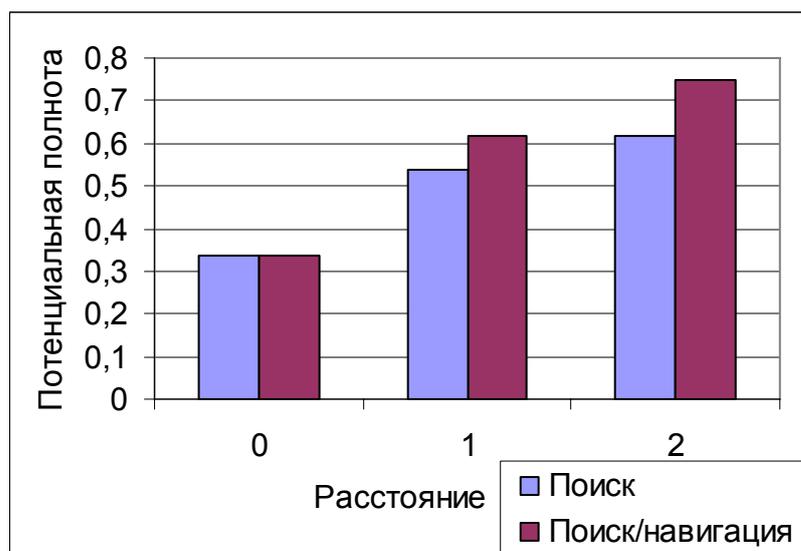


Рис. 5 Зависимость среднего значения потенциальной полноты от расстояния в графе результатов поиска

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1. Разработан поисково-навигационный метод организации доступа к массиву информационных ресурсов, позволяющий повысить эффективность поиска библиотечных ресурсов читателями;
2. Разработаны методы построения ассоциативных связей между разделами каталога и отдельными его записями, являющиеся основой для реализации поисково-навигационного подхода;
3. Предложенный подход реализован в виде подсистемы АБИС библиотеки МГТУ им. Н. Э. Баумана, позволяющей организовать доступ читателей к информационным ресурсам библиотеки. Разработанная подсистема внедрена и функционирует в библиотеке МГТУ им. Н.Э. Баумана;
4. Проведено исследование практической применимости предложенного подхода. Предложены критерии эффективности функционирования поисково-навигационного метода организации доступа к информационным ресурсам, с использованием которых показано улучшение результативности работы пользователей с системой.

РАБОТЫ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Автоматизированная библиотечно-информационная система МГТУ им. Н.Э. Баумана / Г.В. Абрамов, Т.И. Агеева, Б.Г. Трусов, А.Е. Шиваров. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2007. – 50 с.
2. Автоматизированная библиотечно-информационная система технического университета / А.Е. Шиваров, Г.В. Абрамов, О.В. Пескова, Н.А. Белостоцкий

- // Вестник МГТУ им. Н.Э. Баумана. Приборостроение. – 2007. – №4. – С. 21-32.
3. Свидетельство об официальной регистрации программы для ЭВМ №2005610223. Автоматизированная система обслуживания фонда электронных документов / А.Е. Шиваров. – М., 2005. – 1с.
 4. Свидетельство об официальной регистрации программы для ЭВМ №2007610196. Автоматизированная библиотечно-информационная система «Яуза» / А.Е. Шиваров, Г.В. Абрамов, Н.А. Белостоцкий, О.В. Пескова. – М., 2007. – 1с.
 5. Шиваров А. Е. Обслуживание читателей в АБИС библиотеки МГТУ им. Н.Э. Баумана «Яуза»: опыт разработки, внедрения и перспективы развития [Электронный ресурс] // Информационные технологии, компьютерные системы и издательская продукция для библиотек: Материалы 10-й междунар. конф. «LIBCOM-2006». – М., 2006. – Режим доступа: <http://gpntb.ru/libcom6/disk/60.pdf>, свободный.
 6. Шиваров А. Е. Организация процесса обработки, хранения, представления и поиска электронных ресурсов в библиотеке вуза // Научно-техническая информация. Сер. 1. – 2007. – № 2. – С. 6-14.
 7. Шиваров А. Е. Организация фонда электронных документов в библиотеке МГТУ им. Н. Э. Баумана // Культура народов Причерноморья. – 2004. – № 48, т.2. – С. 30-34.
 8. Шиваров А. Е. Подсистема электронного каталога АБИС библиотеки МГТУ им. Н.Э. Баумана «Яуза» [Электронный ресурс] // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: Материалы 14-й междунар. конф. «Крым-2007». – М., 2007. – Режим доступа: <http://gpntb.ru/win/inter-events/crimea2007/cd/122.pdf>, свободный.
 9. Шиваров А. Е. Поисково-навигационный метод организации доступа читателей к информационным ресурсам библиотеки [Электронный ресурс] // Информационные технологии, компьютерные системы и издательская продукция для библиотек: Материалы 11-й междунар. конф. «LIBCOM-2007». – М., 2007. – Режим доступа: <http://gpntb.ru/libcom7/disk/2.pdf>, свободный.
 10. Шиваров А. Е. Полнотекстовый информационный поиск с учетом возможных ошибок оптического распознавания символов // Технологии Microsoft в теории и практике программирования: Труды всероссийской конференции студентов, аспирантов и молодых ученых. – М., 2005. – С. 60.
 11. Шиваров А. Е. Типология ошибок оптического распознавания текста по характеру влияния на процесс индексирования и поиска документов // Технологии Microsoft в теории и практике программирования: Труды всероссийской конференции студентов, аспирантов и молодых ученых. – М., 2006. – С. 74-75.