

Таким образом, мы пришли к наиболее эффективному варианту построения информационной модели системы управления НИРС, не содержащей связей типа «много-ко-многим» и повторяющихся атрибутов у сущностей.

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Иванова Е.В.* Информационная система научно-исследовательской работы студентов // Сб. научных трудов «Наука в решении проблем Верхнекамского промышленного региона», выпуск 6. Березники, 2007.– С. 250-252.

Иванова Елена Владимировна  
Березниковский филиал Пермского государственного технического университета  
E-mail: [loko87@inbox.ru](mailto:loko87@inbox.ru)  
618600, Пермский край, г. Усолье, ул. Набережная, 51-2

Ivanova Elena Vladimirovna  
Berezniki branch of Perm state technical university  
E-mail: [loko87@inbox.ru](mailto:loko87@inbox.ru)  
51-2, Naberezhnaja street, Usolie, 618600, Russia

УДК 004.91

**В.Ю. Баранцов, О.В. Граецкая**

#### **ЭФФЕКТИВНЫЙ ПОИСК В ДОКУМЕНТАЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ**

*Рассмотрены вопросы повышения эффективности деятельности архивных учреждений на основе информационных моделей, позволяющих провести кластеризацию накопившихся документов и разработать систему полнотекстового поиска.*

*Кластерный анализ; дискриминантный анализ; полнотекстовый поиск; иерархическая кластеризация; документальные информационные системы.*

**V.Y. Barancov, O.V. Graetskaya**

#### **EFFICIENT SEARCH IN DOCUMENTARY INFORMATIONAL SYSTEMS**

*The issues to increase the efficiency of archival institutions functioning based on informational models that allow to develop the full-text search system and lead the clustering of accumulated documents are consider.*

*Cluster analysis; discriminant analysis; full-text search; hierarchical clustering; documentary information systems.*

Современное общество накопило колоссальные массивы информации, которые хранятся в специальных учреждениях – архивах. Для использования этой информации необходимо систематизировать накопленные знания. В течении последних нескольких лет происходит заполнение всероссийского каталога архивных фондов, и, хотя работа еще не окончена, уже сегодня обработанную информацию можно использовать для облегчения работы архивистов и исследователей. Однако создаваемый каталог обладает рядом недостатков, главным из которых

является отсутствие автоматизированных систем поиска информации. Кроме того, документы упорядочены по месту расположения в архиве, а данная структура является малоинформативной, т.к. перед людьми встают задачи нахождения документов, содержащих информацию по заданной теме, а они могут находиться в разных местах хранилища. Таким образом, необходимо разработать документальную информационную систему, которая позволит повысить эффективность поиска документов работниками архива и исследователями. Решение данной задачи возможно только с привлечением современных информационных технологий, методов классификации (кластерный и дискриминантный анализ) и внедрения автоматизированных систем поиска информации.

Для того чтобы найти конкретный объект с заданными свойствами, используют следующие методы: индексирование, семантическая навигация (гипертекст), полнотекстовый поиск. Рассмотрим полнотекстовые информационно-поисковые системы (ИПС), использование которых в архивных учреждениях оптимально. В них пользователи освобождаются от необходимости сложной предварительной структуризации предметной области и затратных процедур индексирования при накоплении, получении и агрегировании текстовых документов. Полнотекстовые ИПС строятся на основе информационно-поисковых языков дескрипторного типа, имеется генеральный указатель, в общем виде представляющий перечисление всех слов (словоформ), имеющихся в документах хранилища, с указанием координатного местонахождения каждого слова (№документа – №абзаца – №предложения – №слова). Поисковое пространство отражает весь текст документа, а не только его смысловое содержание.

Таким образом, для достижения поставленной цели необходимо решить следующие задачи:

- 1) провести системный анализ аспектов деятельности архива, связанных с поиском документов;
- 2) выбрать инструмент для проведения дискриминантного и кластерного анализа;
- 3) провести кластерный анализ хранимых документов, позволяющий выделить классы однородных объектов с помощью иерархической кластеризации;
- 4) реализовать процедуру дискриминантного анализа, позволяющую вновь поступающий (или вносимый в базу данных) документ отнести к одному из выделенных классов;
- 5) разработать автоматизированную систему полнотекстового поиска.

Реализация поставленных задач осуществляется на примере государственного учреждения «Центр документации новейшей истории Ростовской области» (ГУ «ЦДНИРО»). В результате выполнения работы проведён системный анализ деятельности ГУ «ЦДНИРО», который выявил недостатки в системе поиска и предоставления доступа к документам (рис. 1). Сегодня процесс поиска документа занимает не менее двух недель и сильно зависит от квалификации и времени работы в данной организации хранителя архива. Если по прошествии некоторого времени документ потребуются ещё раз, то время, потраченное на его поиски, существенно не сократится.

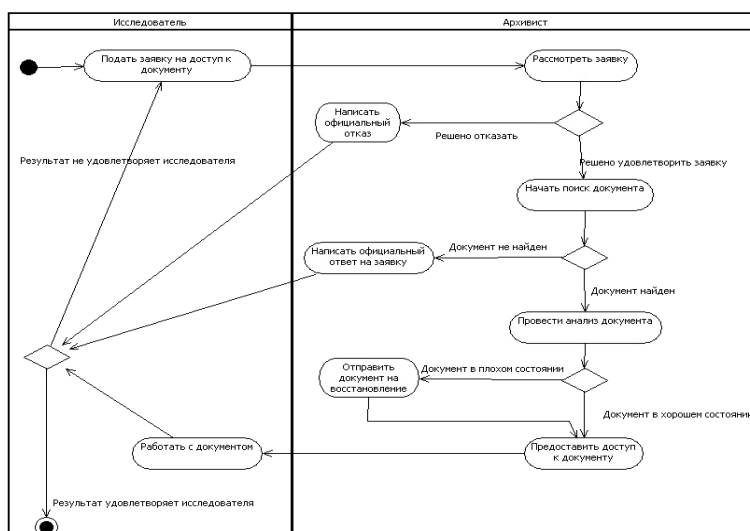


Рис. 1. Диаграмма деятельности «Получение доступа к документу»

Диаграмма на рис. 2 описывает процесс поиска документов с использованием базы данных, при условии, что не все документы были переведены в цифровой вид. Поиск в данном случае начинается с поиска в базе данных, и лишь затем, если документ не найден, он переносится в хранилища. Данный алгоритм позволяет экономить общее время, затрачиваемое на поиск документов, даже при очень незначительном количестве оцифрованных документов. Оцифровка же всех найденных документов позволит как постепенно накапливать информацию в базе данных, так и минимизировать время повторного доступа к документу.

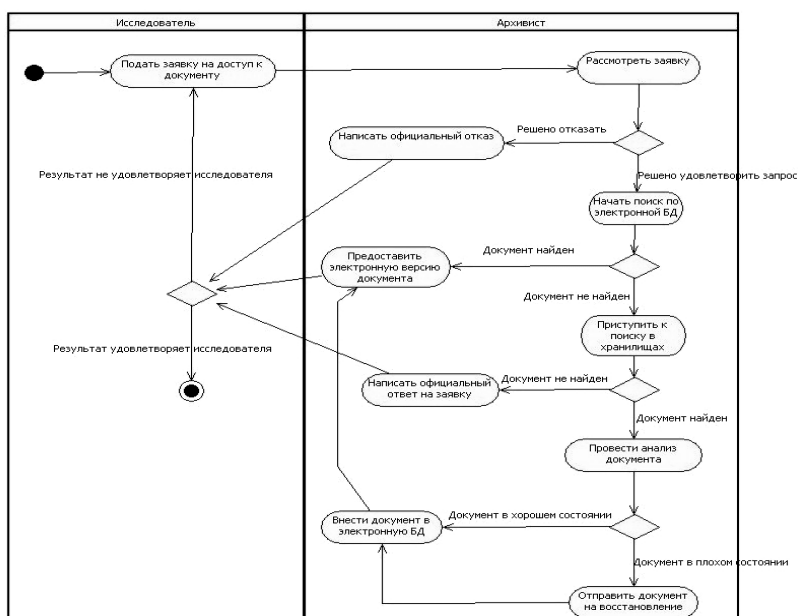


Рис. 2. Диаграмма деятельности «Получение доступа к документу с использованием базы данных»

Кластеризация документов базы данных позволит исследователям самостоятельно искать документы с необходимой им информацией, причём делать это эффективнее, чем архивисты (рис. 3 и 4). Архивист по запросу исследователя ищет конкретный документ, сам же исследователь будет иметь возможность искать некоторую совокупность документов, объединённых общими признаками, постепенно сужая круг поисков, тем самым, отсеивая ненужную для себя информацию.



Рис. 3. Диаграмма деятельности «Поиск документа»

Из всего многообразия методов полнотекстового поиска документов для решаемых в архиве задач больше всего подходят методы на основе пространственно-векторных моделей, которые и будут использованы при разработке автоматизированной системы поиска.

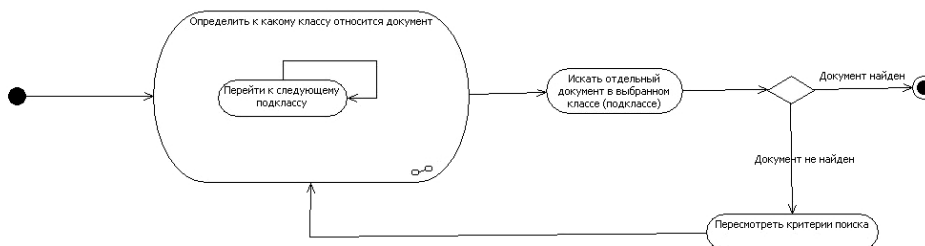


Рис. 4. Диаграмма деятельности «Поиск документа в БД после кластеризации»

Предложенные информационные модели, методы и алгоритмы позволяют повысить эффективность работы архива путём ускорения выполнения работ за счёт внедрения в работу современных технологий. Прделанная работа открывает пути для дальнейшего развития организации и оптимизации её деятельности.

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Ларман К. Применение UML и шаблонов проектирования – М.: Вильямс, 2004. – 620 с.
2. Леоненков. Самоучитель UML. [Электронный ресурс]. – Режим доступа: <http://khpri-ip.mipk.kharkiv.edu/library/case/leon/> (дата обращения: 23.03.2009).

Граецкая Оксана Владимировна  
 Федеральное государственное образовательное учреждение высшего профессионального образования «Южный федеральный университет»  
 E-mail: [g\\_oks@inbox.ru](mailto:g_oks@inbox.ru).  
 344090, Ростов-на-Дону, ул. Мельчакова, 10  
 Тел.: +7(8632)696991

Баранцов Владимир Юрьевич  
Федеральное государственное образовательное учреждение высшего профессионального образования «Южный федеральный университет»

E-mail: [vova\\_barancov@mail.ru](mailto:vova_barancov@mail.ru).

Graetskaya Oksana Vladimirovna  
Federal State-Owned Educational Establishment of Higher Vocational Education «Southern Federal University»

E-mail: [kaf\\_sau@mail.ru](mailto:kaf_sau@mail.ru)

10, Melchikova street, Rostov-on-Don, 344090

Phone: +7(8632)696991

Barancov Vladimir Yurievich  
Federal State-Owned Educational Establishment of Higher Vocational Education «Southern Federal University»

E-mail: [vova\\_barancov@mail.ru](mailto:vova_barancov@mail.ru).

УДК 621.306

**А. А. Строщев, А. Л. Оганесян, М. А. Григорян**

**ТЕОРЕТИКО-ИГРОВАЯ ОПТИМИЗАЦИЯ АЛГОРИТМОВ КОНТРОЛЯ  
СЛОЖНОЙ СИСТЕМЫ НА ОСНОВЕ КЛАССИЧЕСКИХ МОДЕЛЕЙ  
МАТРИЧНЫХ ИГР С ОГРАНИЧЕНИЯМИ-НЕРАВЕНСТВАМИ**

Предложена методика теоретико-игровой оптимизации алгоритмов контроля на основе классических моделей матричных игр с ограничениями-неравенствами. Она позволяет учесть априорные данные об интервалах неопределённости стохастического описания неопределённых факторов.

Теоретико-игровая оптимизация; алгоритмы контроля; матричные игры с ограничениями.

**A.A. Strotsev, A.L. Oganessian, M.A. Grigoryan**

**GAME-THEORETICAL OPTIMIZATION OF CONTROL ALGORITHMS OF  
COMPLICATED SYSTEM BASED ON CLASSICAL MODEL OF MATRIX  
GAMES WITH CONTINGENCIES –INEQUALITIES**

The procedure of game-theoretical optimization of algorithms of control based on classical models of matrix games with contingencies-inequalities was suggested. It is allow to consider the aprioristic date of indeterminacy intervals of stochastic exposition of indefinite factors.

Game-theoretical optimization; algorithms of complicated; matrix games with contingencies.

Эффективность функционирования сложной системы (СС) зависит от качества алгоритмов ее контроля. Методы оптимизации алгоритмов контроля можно классифицировать относительно информационных условий выработки решения, принятых в теории принятия решений: определённости, риска и неопределённости, связанных соответственно с наличием определённых, стохастических и неопределённых факторов.

Нормальный период эксплуатации СС связан с действием, как правило, случайных факторов, имеющих вероятностное описание. Однако периоды прира-