

**Е. А. СОЛОВЬЕВА**, д-р техн. наук, зав. Кафедрой СИ ХНУРЭ,  
**М. Ю. ФРОЛОВ**, аспирант ХНУРЭ

## **ОРГАНИЗАЦИЯ ЛОГИЧЕСКОЙ СТРУКТУРЫ ЭЛЕКТРОННЫХ АРХИВОВ НА ОСНОВЕ ЕСТЕСТВЕННОЙ КЛАССИФИКАЦИИ**

Проводиться аналіз можливості використання системологічного класифікаційного аналізу (СКА) на основі природної класифікації (ПК) у якості основи логічної структури організації електронних архівів для розв'язання деяких проблем організації дослідницького пошуку в електронних архівах.

Проводится анализ возможности использования системологического классификационного анализа (СКА) на основе естественной классификации (ЕК) в качестве основы логической структуры организации электронных архивов для решения некоторых задач организации исследовательского поиска в электронных архивах.

This article examines the possibility of using systemological classification analysis (SCA) on the basis of natural classification (NC) as the basis in the logical structure of the electronic files organization to solve some problems of research in search of electronic archives.

**Постановка проблемы.** Сегодня наблюдается небывалый подъем комплексной компьютеризации предприятий, корпораций и целых отраслей, причем первостепенная роль отводится построению автоматизированных систем документооборота и делопроизводства. Однако с течением времени оказалось, что внедрив мощные системы управления электронными документами, многие корпорации не получили ожидаемого эффекта. В чем причина? Огромное количество - до 80% - оперативной и справочной информации по-прежнему остается на бумажных носителях и складывается в библиотеках или архивах [1]. Ручной поиск и обработка необходимых документов в таких бумажных хранилищах становится узким звеном функционирования многих крупных корпораций. Одним из возможных решений является внедрение и использование электронного документооборота, в том числе электронного архива (ЭА).

Электронные архивы повышают эффективность работы с корпоративными информационными ресурсами, в частности управленческой, финансовой, научно технической и проектно-конструкторской документацией. Эти документы зачастую существуют только в бумажном виде, плохо структурированы, хранятся в разных базах данных, на компьютерах отдельных пользователей и др.

Электронный архив – это единое структурированное хранилище электронных документов, на основе которого создается законченная система документооборота предприятия [2, 3, 4 и др.].

Электронный архив выполняет множество различных функций по работе и управлению документами, но для конечного пользователя принципиально важно исполнение поисковых функций.

В зависимости от объема хранимых документов различают три типа ЭА:

1. Малые архивы - характеризуются наличием небольшого числа хранимых документов (обычно не более 100 тысяч единиц). Общий объем архива до 500 Гб [5].
2. Средние архивы - количество документов от 100 000 до 2 000 000. Общий объем архива от 0,5 до 5 Тб [5].
3. Большие архивы - количество документов в архиве более 2 000 000. Общий объем архива более 5 Тб [5].

Существует две основных вида поиска документов в электронных архивах – атрибутивный и полнотекстовый поиск. Обычно малые архивы содержат атрибутивный поиск - каждому документу присваивается набор определенных атрибутов, присваиваемых документу во время его размещения в архиве. В дальнейшем документ ищется на совпадение значений этих атрибутов полям запроса [6, 7]. К атрибутам документа относят имя документа, время создания, автор, тип документа и т.п. Очевидно, что список таких атрибутов можно расширить. Совокупность атрибутов документа называется карточкой документа [7]. Эти поля могут заполняться произвольно или из predetermined справочников.

Средние и большие типы архивов помимо атрибутивного поиска содержат еще и полнотекстовый - обрабатывается все содержание документа, и затем по любому слову, входящему в данный документ, можно найти сам документ. Существует комбинация полнотекстового и атрибутивного поиска, когда атрибуты документа обрабатываются так же, как все содержание документа [7].

Однако, несмотря на сегодняшнее развитие информационных технологий средств поиска в ЭА, существуют некоторые проблемы, которые и будут рассмотрены и проанализированы в данной статье.

**Анализ последних исследований и публикаций.** Анализ организации поиска в ЭА, проводимый в данной работе, опирается на материалы основных фирм-производителей ЭА АйТи, IBM, IC, Optima, Евфрат и др., содержащие общую информацию о продукте, основные функции, круг решаемых задач, технические характеристики и т.п.

**Формулировка цели статьи.** Целью данной статьи является проведение анализа возможности использования метода системологического классификационного анализа (СКА) на основе естественной классификации (ЕК) [8] для построения логической структуры электронных архивов для решения некоторых проблем организации поиска в ЭА.

**Изложение основного материала.** Существует два подхода к поиску документов. Первый подход состоит в том, что в процессе поиска ищется документ, который точно существует в системе, и задача состоит в том,

чтобы процесс поиска свелся к нахождению требуемого документа или документов – точный поиск. Второй подход состоит в том, что ищутся все документы по интересующему вопросу – исследовательский поиск [7]. Для последнего подхода присуще такое понятие, как шум – часть документов в результирующем наборе, где встретились подобные по синтаксису, но не по смыслу слова [7], т.е. документов нерелевантных поисковому запросу. Существует два основных типа поиска. Атрибутивный - каждому документу присваивается набор определенных атрибутов, присваиваемых документу во время его размещения в архиве [7]. В дальнейшем документ ищется на совпадение значений этих атрибутов полям запроса. Второй тип поиска, носит название полнотекстовый - автоматически обрабатывается все содержание документа, и затем по любому слову, входящему в данный документ, можно найти сам документ [7].

Соответственно существует зависимость между типами поиска и подходами к поиску. Для поиска известного документа, очевидно, применим атрибутивный поиск, для исследовательского поиска - полнотекстовый.

Рассмотрим ситуацию проведения исследовательского поиска для всех типов архивов. Ниже на рис. 1 показана схема организации поиска в малых ЭА.

Как было отмечено выше, для исследовательского поиска атрибутивный не подходит и, следовательно, остается навигация по логической структуре ЭА. Здесь и возникает одна из проблем поиска. Так как в качестве логической структуры выступает иерархическая структура, то при малом количестве уровней это делает понятной навигацию - пользователь представляет, где искать нужный документ. Однако в случае большого количества ветвей иерархии и их глубины эффективность и скорость поиска снижается, так как каждому новому уровню может соответствовать свое основание деления. Одним из возможных способов устранения данных недостатков является использование естественной классификации в качестве логической структуры ЭА, что обеспечит единое основание деления для всей классификации [8].

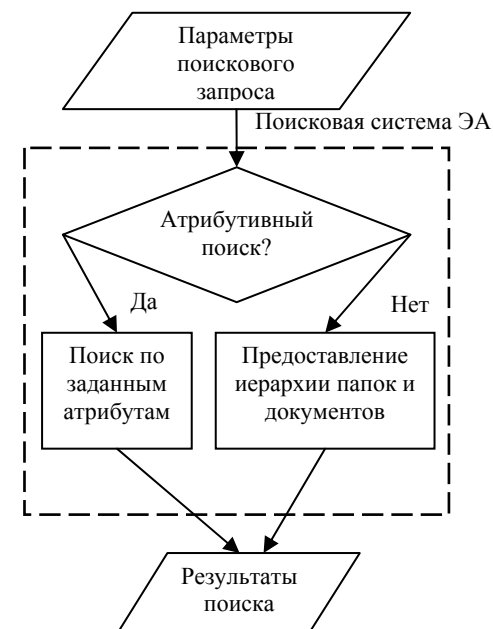


Рис. 1. Схема организации поиска в малых ЭА

В случае со средними и большими ЭА общая схема исследовательского поиска документов выглядит как показано на рис. 2.

Здесь появляется еще одна проблема, связанная с поиском в ЭА. Для достижения эффективного результата при использовании полнотекстового поиска необходимо составить корректный запрос к поисковой системе электронного архива. Но даже при условии отлично составленного запроса конечный результат поиска будет содержать так называемый «шум», т.е. документы нерелевантные запросу. Это происходит отчасти из-за того, что система поиска просматривает всю базу данных индексов ЭА. Используя метод системологического классификационного анализа для построения параметрической классификации в качестве логической структуры, мы получим для каждого документа свое главное отличительное функциональное свойство [8]. При исследовательском поиске функциональное свойство, по сути, является целью данного поиска. Пользователь, задав его в параметры запроса, указывает необходимую ветвь иерархии для поиска. Данную ветвь можно использовать для дополнительного отсеивания документов, полученных в результирующем наборе, оставляя только те документы, которые находятся в указанной ветви иерархии, что также сможет снизить уровень шума в конечном результате.

данными, электронный архив, документооборот *Алексей Лихачев* «САПР и графика» 3'2001 [Электронный ресурс]–Режим доступа Интернет:<http://www.sapr.ru/article.aspx?id=6971&iid=285>. 7. Концепция построения систем автоматизации документооборота, *Анатолий Гавердовский*, АО ВЕСТЬ, Открытые системы. #1/97, стр. 29-34. 8. *Соловьева Е.А.* Естественная классификация: системологические основания - Харьков: ХНУРЕ, 1999 - 222 с.

Поступила в редколлегию 15.12.09

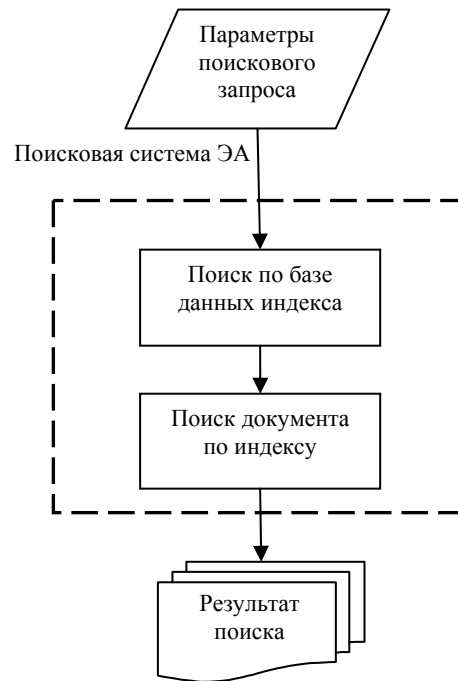


Рис. 2. Схема исследовательского поиска в средних и больших ЭА

**Выводы.** При использовании естественной классификации в качестве логической структуры ЭА появляется возможность частично решить некоторые проблемы, связанные с исследовательским поиском в электронных архивах: снижение шума в конечном результирующем наборе файлов в больших и средних архивах, а также повышение эффективности исследовательского поиска в малых архивах при развитой структуре цепочек документов.

**Список литературы:** 1. «Концепция построения электронных архивов», *Алексей Марков* [Электронный ресурс]. – Режим доступа Интернет: [http://www.alecte-archiv.ru/page.jsp?pk=node\\_1129899705713](http://www.alecte-archiv.ru/page.jsp?pk=node_1129899705713). 2. Электронный архив [Электронный ресурс]. – Режим доступа Интернет: <http://www.scanit.ru/archive/elar/>. 3. Концепция электронного архива [Электронный ресурс] – Режим доступа Интернет: <http://www.mosarchiv.com/sub/service/useful/article/13>. 4. Обзор и анализ основных систем автоматизации документооборота, *Фролов М.Ю.* Системы обработки информации, Харьков, 2009 – 131-134 стр. 5. Электронные архивы [Электронный ресурс]. – Режим доступа Интернет: <http://www.korusecm.ru/solutions/elib/>. 6. ProjectWise как система управления проектными