

# ОБ ОДНОЙ МОДЕЛИ ОПТИМИЗАЦИИ ДОКУМЕНТОПОТОКОВ, РЕАЛИЗУЕМОЙ ПРИ СОЗДАНИИ СИСТЕМЫ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА\*

А. М. Гудов, С. Ю. ЗАВОЗКИН

*Кемеровский государственный университет, Россия*

e-mail: good@kemsu.ru, shade@kemsu.ru

This article addresses one of the approaches to formulation of the problem of the document flow optimization according to a certain specified criterion. The approach is based on searching out the optimal document propagation paths, which are determined by the architecture design and the structure of online document circulation system. The formulation of the problem relies on the analysis of the existing structure and established business-processes. Solution of the problem was obtained empirically in accordance with the results of monitoring of the system characteristics that enabled us to select the optimal system architecture.

## Введение

Система электронного документооборота (СЭД) создается для решения проблем автоматизации хранения, поиска и обработки больших массивов информации, а также обеспечения безопасности ее хранения, передачи и возможности совместного использования. Еще одна ключевая задача СЭД, которую она может решать, — это объединение всех подсистем в одну информационную среду на основе определения единого понятия документа и интерфейса обмена документами между подсистемами [1].

Реализуемая СЭД должна не просто автоматизировать движение документов, но и делать это по возможности наиболее эффективно с точки зрения выбранного критерия. Это, в свою очередь, приводит к необходимости решения задачи оптимизации документопотоков на основе методов математического моделирования — нахождения оптимальных по определенному критерию путей прохождения документов, определенных архитектурой и структурой системы. Решение данной проблемы должно основываться на всестороннем анализе принятых бизнес-процессов высшего учебного заведения.

В настоящей статье рассмотрен один из подходов к постановке задачи оптимизации документопотоков, автоматизируемых системой электронного документооборота, которая в данный момент реализуется в Кемеровском государственном университете.

---

\*Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации (грант № 4828).

© Институт вычислительных технологий Сибирского отделения Российской академии наук, 2006.

## 1. Общие положения

На основе результатов анализа существующего документооборота вуза, доступного описания систем электронного документооборота, а также сведений из литературы, посвященных вопросам построения автоматизированных информационных систем и интеграции их компонентов были разработаны требования к СЭД и определено необходимое множество базовых функций системы [2, 3]: создание, хранение, редактирование, удаление документов; управление доступом к документам; доставка документов пользователям по заранее определенному или динамическому маршрутам; ведение журнала регистрации документов; выполнение и контроль над выполнением заданий, связанных с документооборотом; обеспечение интеграции с другими информационными системами и в первую очередь с подсистемами интегрированной аналитической информационной системы управления вузом; максимальная ориентация на целевую аудиторию; наличие удобного web-интерфейса.

В поле компетенции СЭД попадают сами электронные документы, процесс движения электронных документов как внутри вуза, так и между СЭД и информационными системами, а также режим управления электронными документами.

С учетом рекомендаций для представления в информационной системе электронного документа дано его определение как информационной пары, представляющей собой описание (метаданные) документа, позволяющее его однозначно идентифицировать, и содержимое (тело) документа [3–5]. Удобство работы с такой “моделью” электронного документа ускоряет время доступа к информации, время поиска данных в СЭД, а также уменьшает объем трафика, циркулирующего в коммуникационной среде. Помимо этого, включение в метаданные электронного документа информации о его жизненном цикле позволяет реализовать эффективное управление движением документов.

Наряду с определением электронных документов и их структуры рассмотрено понятие маршрута следования документа [3]. Все маршруты следования документов в системе подразделяются на две группы: стационарные (не изменяемые в процессе жизненного цикла документа) и динамические. Маршруты следования, определенные для каждого документа, моделируют в СЭД документопотоки, сложившиеся в бизнес-процессах вуза. Следовательно, можно предположить, что оптимизация маршрутов обмена документов наряду с возможностью приоритетной доставки по маршрутам следования метаданных позволит минимизировать затраты на обслуживание документопотоков в СЭД.

Эти допущения положены далее в основу построения задачи оптимизации документопотоков.

## 2. Оптимизация документопотоков СЭД

Математическая модель оптимизации документопотоков построена с использованием методики, использованной в работе [6]. В ее основе лежит “структурный” подход, при котором учитывается конкретная структура исследуемой организации и оптимизируются документопотоки в рамках этой заданной структуры. Такой подход имеет ряд недостатков, в первую очередь это необходимость “учета” в получаемой целевой функции структуры организации при ее изменении.

При построении математической модели оптимизации авторами предложен “архитектурный” подход — оптимизация документопотоков на основе анализа возможного использования различных архитектурных решений при проектировании системы электронного документооборота.

В общем случае задача оптимизации документопотоков на основе методов математического моделирования состоит в нахождении параметров СЭД, обеспечивающих оптимальное движение документов в соответствии с выбранными критериями. Часто выделяют следующие основные объекты оптимизации:

- объем занимаемого документами места в хранилище СЭД, включая все копии документа;
- время, затрачиваемое на получение документа из хранилища по запросу пользователя;
- время, затрачиваемое на получение документа из некоторой информационной системы.

Рассмотрим варианты решений по оптимизации документопотоков.

**Первый вариант** состоит в минимизации времени получения требуемого электронного документа пользователем. Оно складывается из времени, затрачиваемого на передачу запроса, времени поиска (и в случае необходимости формирования) запрашиваемого документа, а также времени его непосредственной передачи пользователю. При таком подходе необходимо оценивать параметры коммуникационной сети (например, предельную пропускную способность каналов передачи данных, архитектуру сети и т. д.), серверов, на которых расположены как само приложение СЭД, так и хранилища информации (например, тактовая частота процессора, объем оперативной памяти, объем и скорость доступа к запоминающим устройствам и т. д.). В этом случае целевую функцию  $F_1$  можно определить следующим образом:

$$F_1 = t' + t'' \rightarrow \min, \quad (1)$$

где  $t'$  — среднее время получения пользователем электронного документа из хранилищ СЭД;  $t''$  — среднее время получения пользователем электронного документа из информационных систем. Очевидно, что каждый из этих параметров будет зависеть от архитектуры распределенного хранилища (хранилищ) документов и как следствие архитектуры информационных систем, использующих данные из этих хранилищ.

**Второй вариант** предполагает минимизацию стоимости передачи документов между пользователями и хранилищами СЭД, а также между пользователями и другими информационными системами. В этом случае целевую функцию  $F_2$  можно представить как

$$F_2 = \sum_{i=1}^N \sum_{j=1}^{n_i} W'_{ij} + \sum_{i=1}^M \sum_{j=1}^{m_i} W''_{ij} \rightarrow \min, \quad (2)$$

где  $n_i$  — количество документов в  $i$ -м хранилище  $H_i$ ;  $m_i$  — количество документов в  $i$ -й информационной системе  $I_i$ ;  $N$  — количество хранилищ;  $M$  — количество информационных систем, с которыми обменивается информацией СЭД;  $W'_{ij}$  — стоимость передачи документа  $d_j$  между пользователем и  $H_i$ ;  $W''_{ij}$  — стоимость передачи документа  $d_j$  между пользователем и  $I_i$ . И в этом случае стоимости как параметры функции будут определяться архитектурой распределенной системы.

**Третий вариант** состоит в минимизации объемов электронных документов, хранящихся в СЭД. Оптимизация по данному признаку ведет к снижению неперiodических затрат на хранение электронных документов, а также на их обновление, поскольку минимизация дублирования электронных документов приводит к снижению затрат на внесение изменений во все хранилища информации. При этом значительно увеличивается быстродействие поиска электронного документа. Однако уменьшение объемов информации за счет ликвидации ее дублирования имеет и отрицательные стороны. Это в первую очередь

возможность потери информации при неосторожном удалении электронного документа, а также увеличения стоимости передачи за счет необходимости передачи всех нужных документов между хранилищами СЭД. В данном случае целевая функция  $F_3$  будет выглядеть следующим образом:

$$F_3 = \sum_{i=1}^N \sum_{j=1}^{n_i} v_{ij} \rightarrow \min, \quad (3)$$

где  $v_{ij}$  — объем электронного документа  $d_j$  в  $H_i$ ;  $n_i$  — количество документов в  $H_i$ ;  $N$  — количество хранилищ СЭД. Такой подход напрямую не зависит от архитектуры системы (распределенное хранилище рассматривается как отдельные объекты системы), но имеет ряд сильных ограничений, которые не учитывают “окружение” будущей системы.

**Четвертый вариант** состоит в минимизации затрат на обновление электронных документов. При высокой степени изменчивости информации этот вариант приводит к значительному уменьшению количества копий. Но следует отметить, что уменьшение объемов хранимой информации за счет ликвидации ее дублирования имеет отрицательные стороны. Так, увеличение документопотоков информации, передаваемой между различными хранилищами СЭД, вызывает увеличение времени исполнения операций и стоимости поиска информации из-за необходимости обращения именно к тем хранилищам, где находятся запрашиваемые документы. Однако если хранилища информации территориально разделены, то время и стоимость передачи могут значительно превысить экономию от уменьшения объемов хранимой информации. Кроме того, при этом варианте значительно снижены показатели надежности системы. Данный подход позволяет построить целевую функцию  $F_4$ :

$$F_4 = \sum_{i=1}^N \sum_{j=1}^{n_i} u_{ij} \rightarrow \min, \quad (4)$$

где  $u_{ij}$  — стоимость обновления  $d_j$  в  $H_i$ ;  $n_i$  — количество документов в  $H_i$ ;  $N$  — количество хранилищ СЭД.

Стоимость модификации электронного документа включает стоимость получения нужного документа пользователем, стоимость передачи измененной информации и стоимость непосредственно изменения содержимого документа. При выборе сервера, предназначенного для работы с большими объемами данных, стоимость непосредственного обновления информации существенно меньше стоимости поиска нужного электронного документа в базе и его передачи. Таким образом, нет смысла рассматривать отдельно задачу минимизации стоимости модификации электронного документа — она решается в рамках второго варианта.

Подводя итог обсуждению вышеперечисленных вариантов оптимизации, можно сделать вывод, что ни один из них не может быть применен в “чистом виде”. Оптимальность документопотоков может быть достигнута только на основе интегрального пути — сочетания первых трех вариантов.

Воспользуемся хорошо зарекомендовавшим себя экономическим критерием [6], который минимизирует сумму затрат на хранение, передачу, поиск, обновление электронных документов и обеспечение безопасности хранения и передачи информации. Следует отметить, что наряду со стоимостью хранения и передачи информации существует другой важный для СЭД критерий — время передачи информации. Его можно выразить через стоимость передачи документов пользователю или введением дополнительного ограничения на максимально возможный временной интервал, в течение которого пользова-

тель может ожидать документ по своему запросу. Мы в дальнейшем объединим оба эти подхода.

Таким образом, интегральный экономический критерий  $S_{\text{инт}}$  можно определить как

$$S_{\text{инт}} = (S_{\text{хр}} + S_{\text{пхр}} + S_{\text{пиф}}) \rightarrow \min, \quad (5)$$

где  $S_{\text{хр}}$  — стоимость хранения информации;  $S_{\text{пиф}}$  — стоимость получения данных из информационных систем;  $S_{\text{пхр}}$  — стоимость получения данных пользователем из хранилищ СЭД.

Сформулируем постановку задачи оптимизации передачи данных в соответствии с определенным критерием в терминах метода целочисленного программирования [7, 8]. Рассмотрим отдельно каждый из объектов оптимизации и модифицируем его оценку с учетом возможных архитектурных решений для проектируемой системы.

## 2.1. Объем занимаемого электронными документами места в хранилищах СЭД

Введем индикатор, показывающий, какие из электронных документов находятся в хранилищах СЭД. Определим  $x_{ij}$  следующим образом:

$$x_{ij} = \begin{cases} 1, & d_j \in H_i, \\ 0, & d_j \notin H_i, \end{cases} \quad (6)$$

где значение, равное “1”, показывает, что документ  $d_j$  хранится в  $i$ -м информационном хранилище  $H_i$ ,  $j = \overline{1, n_i}$ ,  $n_i$  — количество электронных документов в  $H_i$ ,  $i = \overline{1, N}$ ,  $N$  — количество хранилищ.

Каждый электронный документ  $d_j$  может иметь определенное количество копий  $r_j$  в различных хранилищах:

$$\sum_{i=1}^N x_{ij} = r_j. \quad (7)$$

Использование выражения (7) в качестве ограничения создает возможность управления избыточностью данных в СЭД.

Возможность хранения документов ограничена объемом хранилищ информации:

$$\sum_{i=1}^N \sum_{j=1}^{n_i} x_{ij} v_{ij} \leq \sum_{i=1}^n O_i. \quad (8)$$

Здесь  $v_{ij}$  — объем  $d_j$  в  $H_i$ ;  $n_i$  — количество документов в  $H_i$ ;  $N$  — количество хранилищ;  $O_i$  — объем доступной памяти для размещения электронного документа в  $H_i$ .

Так как документ состоит из двух частей — содержимого и метаданных, введем индикаторы  $\text{met}_{ijk}$  и  $\text{con}_{ijk}$ , показывающие, из каких метаданных и какого содержимого состоит  $d_j$ :

$$\text{met}_{ijk} = \begin{cases} 1, & M_j^k \in H_i, \\ 0, & M_j^k \notin H_i; \end{cases} \quad (9)$$

$$\text{con}_{ijk} = \begin{cases} 1, & C_j^k \in H_i, \\ 0, & C_j^k \notin H_i, \end{cases} \quad (10)$$

где  $M_j^k$  обозначаем  $k$ -е метадачное  $d_j$ , а  $C_j^k$  —  $k$ -е содержимое  $d_j$ ,  $d_j = \{M_j^k, C_j^k\}$ . Тогда общий объем документа представляется как

$$v_{ij} = \sum_{k=1}^{n'_{ij}} \text{met}_{ijk} v'_{ijk} + \sum_{k=1}^{n''_{ij}} \text{con}_{ijk} v''_{ijk}. \quad (11)$$

Здесь  $v'_{ijk}$  — объем  $M_j^k$  электронного документа  $d_j$  в  $H_i$ ;  $v''_{ijk}$  — объем  $C_j^k$  электронного документа  $d_j$  в  $H_i$ ;  $n'_{ij}$  — число метадачных  $d_j$  в  $H_i$ ;  $n''_{ij}$  — число компонентов коллекции, представляющих содержимое  $d_j$  в  $H_i$ . Тогда формулу (8) можно переписать как

$$\sum_{k=1}^{n'_{ij}} \text{met}_{ijk} v'_{ijk} + \sum_{k=1}^{n''_{ij}} \text{con}_{ijk} v''_{ijk} \leq \sum_{i=1}^n O_i. \quad (12)$$

Данное ограничение регулирует общий объем электронных хранилищ документов и зависит только от двух параметров — физического объема внешней памяти и архитектуры распределенной системы хранилищ. Ограничение не влияет непосредственно на оптимизацию документопотоков, однако может оказаться довольно важным фактором для оценки количества хранилищ в распределенной системе и необходимой избыточности хранения информации.

## 2.2. Время получения электронного документа из $i$ -го хранилища СЭД

Время получения электронного документа из  $H_i$  складывается из времени, затрачиваемого на передачу запроса в  $H_i$ , времени, затрачиваемого на поиск электронного документа в  $H_i$ , и времени, затрачиваемого на передачу документа из  $H_i$  пользователю.

Период, характеризующий среднее время ожидания электронного документа по запросу пользователя из  $H_i$ , обозначим через  $t_i$ . При этом необходимо отметить, что существует максимально допустимое время ожидания —  $T_i$ , которое не может быть превышено. Таким образом, можно сформулировать следующее ограничение:

$$t_i \leq T_i. \quad (13)$$

Частоту запросов к  $H_i$  обозначим  $Q_i$ . Очевидно, что частота обращений  $Q_i$  влияет на эффективность распределения информации в системе, поэтому для включения этого параметра в модель выразим  $Q_i$  через  $x_{ij}$ :

$$Q_i = \sum_{j=1}^{n_i} \eta_{ij} x_{ij}, \quad (14)$$

где  $\eta_{ij}$  — частота обращений к  $d_j$  в  $H_i$  за единицу времени,  $n_i$  — количество электронных документов в  $H_i$ .

Следует отметить, что время  $t_i$  зависит от  $v_{ij}$  и  $Q_i$ . Обозначим пропускную способность канала связи, по которому передаются электронные документы между пользователем и хранилищем, через  $R_i$ . Тогда среднее время передачи документа можно представить в следующем виде:

$$t_i = \frac{1}{Q_i} \sum_{j=1}^{n_i} t_{ij} \eta_{ij} x_{ij}, \quad (15)$$

где  $t_{ij} = v_{ij}/R_i$  — время, требуемое на передачу электронного документа  $d_j$  объемом  $v_{ij}$  из  $H_i$ .

Подставляя (15) в (13), получаем ограничение

$$\frac{1}{Q_i} \sum_{j=1}^{n_i} t_{ij} \eta_{ij} x_{ij} \leq T_i. \quad (16)$$

Через это ограничение можно выразить максимальный объем электронных документов, передаваемых между пользователем и хранилищами во временной период  $T_i$ . Особенно это становится необходимым при анализе топологии сети передачи данных как платформы для организации документопотоков. Увеличение объема передаваемой информации системой документооборота непосредственно будет влиять на выбор архитектурного решения для построения системы, а возможно, и для модификации корпоративной сети организации.

### 2.3. Время ожидания пользователем электронного документа из “внешней” информационной системы

В настоящее время при создании систем электронного документооборота все чаще делается упор на тот факт, что такая система должна не только обладать “традиционным” функционалом, но и служить платформой для интеграции работы с другими информационными системами, уже внедренными в рассматриваемой организации. Для учета этого фактора при построении модели введем дополнительное ограничение на время получения документа из “внешней” информационной системы.

Время получения документов из внешней (по отношению к СЭД) системы можно рассмотреть по аналогии со временем получения документов из хранилища. Отличие заключается в значительно больших временных затратах на поиск (а при необходимости и формирование) запрашиваемой информации, а также дополнительном времени, необходимом для регистрации полученного из информационной системы документа в СЭД. Время передачи документа из внешних информационных систем увеличится за счет существенно более низкой пропускной способности канала по сравнению с каналом между пользователем и хранилищем СЭД.

Под регистрацией электронного документа в СЭД понимаем формирование метаданных полученного электронного документа и помещение их в хранилище СЭД. Так как размер метаданных мал по сравнению с размером содержимого электронного документа, время регистрации электронного документа в СЭД также мало по сравнению с остальными временными составляющими интервала ожидания электронного документа из информационной системы. Поэтому временем регистрации документа в СЭД в данном случае можно пренебречь.

Время получения пользователем документа  $d_j$  из информационной системы  $I_i$  можно рассчитать по аналогии со временем получения документа  $d_j$  из хранилища  $H_i$  (15). Введем индикатор  $y_{ij}$ , показывающий, что документ  $d_j$  хранится в информационной системе  $I_i$ :

$$y_{ij} = \begin{cases} 1, & d_j \in I_i, \\ 0, & d_j \notin I_i, \end{cases} \quad (17)$$

где  $j = \overline{1, m_i}$ ,  $m_i$  — общее количество документов в  $I_i$ ,  $i = \overline{1, M}$ ,  $M$  — число информационных систем.

Период, характеризующий среднее время ожидания  $d_j$  по запросу пользователя из  $I_i$ , обозначим через  $\tau_i$ . Максимально допустимое время ожидания пользователем  $d_j$  из  $I_i$  —  $T'_i$ . Соответственно сформулируем следующее ограничение:

$$\tau_i \leq T'_i. \quad (18)$$

Частоту запросов к  $I_i$  обозначим  $Q'_i$ . Выразим  $Q'_i$  через  $y_{ij}$  по аналогии с (14):

$$Q'_i = \sum_{j=1}^{m_i} \eta'_{ij} y_{ij}, \quad (19)$$

где  $\eta'_{ij}$  — частота обращения к  $d_j$  в  $I_i$  за единицу времени;  $m_i$  — количество электронных документов в  $I_i$ .

Обозначим пропускную способность канала связи, по которому передаются электронные документы из  $I_i$ , через  $R'_i$ . Тогда среднее время передачи можно представить в виде

$$\tau_i = \frac{1}{Q'_i} \sum_{j=1}^{m_i} \tau_{ij} \eta'_{ij} y_{ij}, \quad (20)$$

где  $\tau_{ij} = v'_{ij}/R'_i$  — время, требуемое на передачу  $d_j$  объемом  $v'_{ij}$  из  $I_i$ .

Подставляя (20) в (18), получаем

$$\frac{1}{Q'_i} \sum_{j=1}^{m_i} \tau_{ij} \eta'_{ij} y_{ij} \leq T'_i. \quad (21)$$

Смысл этого ограничения такой же, как и в подразд. 2.2. Его также удобнее использовать для получения максимального объема документов, передаваемых из внешних систем в период времени  $T'_i$ .

## 2.4. Определение стоимостей

Выразим стоимость хранения документов в хранилищах СЭД через  $x_{ij}$ :

$$S_{\text{xp}} = \sum_{i=1}^N \sum_{j=1}^{n_i} s_i^{\text{xp}} x_{ij} \left( \sum_{k=1}^{n'_{ij}} \text{met}_{ijk} v'_{ijk} + \sum_{k=1}^{n''_{ij}} \text{con}_{ijk} v''_{ijk} \right), \quad (22)$$

где  $s_i^{\text{xp}} = W_i^{\text{xp}}/V_i^{\text{xp}}$  — стоимость хранения единицы информации в  $H_i$  на определенный период времени (по умолчанию — один год);  $W_i^{\text{xp}}$  — стоимость  $H_i$ ;  $W_i^{\text{xp}} = W_i^{\text{bo}} + W_i^{\text{oo}}$ ,  $W_i^{\text{bo}}$  — стоимость владения оборудованием;  $W_i^{\text{bo}} = W_i^{\text{o}} r d$ ,  $W_i^{\text{o}}$  — стоимость самого оборудования;  $r$  — амортизационный коэффициент;  $d$  — дисконтные ставки (т.е. учет обесценивания);  $W_i^{\text{oo}}$  — стоимость обслуживания оборудования;  $W_i^{\text{xp}} = W_i^{\text{zn}} + W_i^{\text{tn}} + W_i^{\text{ln}} + W_i^{\text{a}}$ ,  $W_i^{\text{zn}}$  — заработная плата администратора,  $W_i^{\text{tn}}$  — стоимость технической поддержки,  $W_i^{\text{ln}}$  — стоимость лицензионной политики,  $W_i^{\text{a}}$  — стоимость аренды;  $V_i^{\text{xp}}$  — объем  $H_i$ .

Определим стоимость получения данных пользователем из хранилища  $H_i$ :

$$S_{\text{пхр}} = \sum_{i=1}^N \left( \frac{1}{Q_i} \sum_{j=1}^{n_i} s_i^{\text{пхр}} t_{ij} \eta_{ij} x_{ij} \right). \quad (23)$$



Здесь  $s_i^{\text{пхр}}$  — стоимость получения единицы информации из  $H_i$ ;  $s_i^{\text{пхр}} = (W_i^{\text{пхр}}/V_i^{\text{пхр}})R_i$ ,  $W_i^{\text{пхр}} = W_i^{\text{во}} + W_i^{\text{т}}$ ,  $W_i^{\text{во}}$  — стоимость владения коммуникационным оборудованием на участке сети от пользователей до  $H_i$ ,  $W_i^{\text{во}} = W_i^{\text{орд}}$ ,  $W_i^{\text{о}}$  — стоимость самого коммуникационного оборудования на участке сети от пользователей до  $H_i$ ;  $W_i^{\text{т}}$  — тарифный план, определяющий стоимость получения информации из  $H_i$ ;  $V_i^{\text{пхр}}$  — объем переданных за выбранный интервал времени документов из  $H_i$ ;  $R_i$  — пропускная способность канала до  $H_i$ .

Определим стоимость получения документов из информационных систем:

$$S_{\text{пиф}} = \sum_{i=1}^M \left( \frac{1}{Q_i} \sum_{j=1}^{m_i} s_i^{\text{пиф}} \tau_{ij} \eta'_{ij} y_{ij} \right), \quad (24)$$

где  $s_i^{\text{пиф}}$  — стоимость получения единицы информации от  $I_i$ ;  $s_i^{\text{пиф}} = (W_i^{\text{пиф}}/V_i^{\text{пиф}})R'_i$ ,  $W_i^{\text{пиф}} = W_i^{\text{во}} + W_i^{\text{т}} + W_i^{\text{по}}$ ,  $W_i^{\text{во}}$  — стоимость владения коммуникационным оборудованием на участке сети от пользователей до  $I_i$ ;  $W_i^{\text{во}} = W_i^{\text{орд}}$ ,  $W_i^{\text{о}}$  — стоимость самого коммуникационного оборудования на участке сети от пользователей до  $I_i$ ;  $W_i^{\text{т}}$  — тарифный план провайдера, определяющий стоимость получения информации из  $I_i$ ;  $W_i^{\text{по}}$  — стоимость владения информационной системой  $I_i$ ;  $V_i^{\text{пиф}}$  — объем переданных за выбранный интервал времени документов из  $I_i$  (по умолчанию — один год);  $R'_i$  — пропускная способность канала до  $I_i$ .

## 2.5. Постановка задачи оптимизации

Для возможности управления вкладом каждой стоимости в итоговое значение целевой функции для нашей задачи необходимо в результирующую целевую функцию внести весовые коэффициенты, а именно:

$$f(X, Y, V, V') = \omega_1 S_{\text{хр}} + \omega_2 S_{\text{пхр}} + \omega_3 S_{\text{пиф}}; \quad (25)$$

$$\sum_{i=1}^3 \omega_i = 1. \quad (26)$$

Окончательный вид целевой функции

$$\begin{aligned} f(X, Y, V, V') = & \omega_1 \sum_{i=1}^N \sum_{j=1}^{n_i} s_i^{\text{хр}} x_{ij} \left( \sum_{k=1}^{n'_{ij}} \text{met}_{ijk} v'_{ijk} + \sum_{k=1}^{n''_{ij}} \text{con}_{ijk} v''_{ijk} \right) + \\ & + \omega_2 \sum_{i=1}^M \left( \frac{1}{Q_i} \sum_{j=1}^{m_i} s_i^{\text{пхр}} t_{ij} \eta_{ij} x_{ij} \right) + \omega_3 \sum_{i=1}^M \left( \frac{1}{Q_i} \sum_{j=1}^{m_i} s_i^{\text{пиф}} \tau_{ij} \eta'_{ij} y_{ij} \right). \end{aligned} \quad (27)$$

Матрицы  $X = \{x_{ij} \ i = \overline{1, n} \ j = \overline{1, n_N}\}$  и  $Y = \{y_{ij} \ i = \overline{1, m} \ j = \overline{1, m_M}\}$  вместе с определяемыми элементами этих матриц объемами  $V = \{v_{ij} \ \forall x_{ij} \neq 0, i = \overline{1, n}, j = \overline{1, n_n}\}$  и  $V' = \{v_{ij} \ \forall y_{ij} \neq 0, i = \overline{1, m}, j = \overline{1, m_M}\}$  составляют набор переменных задачи.

Полностью постановка задачи оптимизации выглядит следующим образом:

$$f(X, Y, V, V') \rightarrow \min \quad (28)$$

при ограничениях

$$\sum_{i=1}^N x_{ij} = r_j; \quad (29)$$

$$\sum_{i=1}^N \sum_{j=1}^{n_i} x_{ij} \left( \sum_{k=1}^{n'_{ij}} \text{met}_{ijk} v'_{ijk} + \sum_{k=1}^{n''_{ij}} \text{con}_{ijk} v''_{ijk} \right) \leq \sum_{i=1}^n O_i; \quad (30)$$

$$\sum_{i=1}^N \left( \frac{1}{Q_i} \sum_{j=1}^{n_i} t_{ij} \eta_{ij} x_{ij} \right) + \sum_{i=1}^M \left( \frac{1}{Q'_i} \sum_{j=1}^{m_i} \tau_{ij} \eta'_{ij} y_{ij} \right) \leq T; \quad (31)$$

$$\sum_{i=1}^3 \omega_i = 1. \quad (32)$$

Для решения задачи оптимизации необходимо найти такие коэффициенты выражения (27), при которых значение  $f(X, Y, V, V')$  становится минимальным для данного набора значений матриц  $X$  и  $Y$  при ограничениях объема памяти, числа копий и времени получения электронных документов, задаваемых выражениями (29)–(32).

## 2.6. Эффективные параметры задачи

Переменными в нашей задаче являются массивы индикаторов  $X$  и  $Y$ , а также определяемые ими объемы  $V$  и  $V'$ . К параметрам уравнения, оптимальные значения которых мы будем искать, относятся:

- число хранилищ СЭД и число внешних информационных систем;
- пропускная способность каналов между пользователями, хранилищами и внешними информационными системами;
- стоимость хранения единицы информации в  $H_i$ ;
- стоимость получения единицы информации из  $H_i$  и  $I_i$ .

Такие параметры, как пропускная способность каналов и объем хранилищ, зависят только от ресурсов вычислительной системы, которая является платформой для развертывания системы документооборота, и серверов, на которых обслуживаются сторонние информационные системы. Варьирование этих параметров возможно только в случае технической модернизации элементов вычислительно-коммуникационной системы, а следовательно, они не зависят от состояния СЭД. С другой стороны, выбор этих параметров будет определяться на этапе архитектурного проектирования системы.

## 2.7. Анализ целевой функции

Рассмотрим подробно каждое из слагаемых целевой функции.

**Стоимость хранения документов в хранилищах СЭД.** При рассмотрении полученной формулы для стоимости хранения документов (23) можно выдвинуть следующие предположения.

1. Если каждый документ представляет собой пару метаданные — содержимое документа, то  $n'_{ij} \leq n''_{ij}$ .

2. Необходимо учитывать, что у “простых” документов (например, одно-, двухстраничных текстовых документов) объем метаданных сравним с объемом содержимого документа. Поэтому для оценки таких документов нельзя пренебрегать членом  $\sum_{k=1}^{n'_{ij}} \text{met}_{ijk} v'_{ijk}$ .

Однако объем документов, подготовленных в самой распространенной офисной системе MS Office, достаточно велик по сравнению с метаданными (например, с их обязательным

набором, составляющим порядка 20–30 метаданных). Исследования документооборота в КемГУ показали, что средний объем документа даже в случае “простых” документов составляет 800...1000 кбайт, в то время как объем метаданных (текстовые поля в базах данных) около 80...150 кбайт. Таким образом, можно сделать предположение об определяющем влиянии на  $S_{xp}$  объема выбранных документов  $v_{ij}$  при  $x_{ij} \neq 0$ . Этот объем определяется в момент выбора пользователем или другими особенностями сценария документооборота.

3. Объем информации, извлекаемой из хранилища  $H_i$  в некоторый небольшой период времени, будет существенно влиять на время передачи этой информации через каналы связи. Задержка, связанная с поиском и извлечением необходимого документа, в современных промышленных (даже реляционных) базах данных незначительна. Уменьшения этого времени можно добиться простыми приемами индексирования и кластеризации данных на уровне СУБД, управляющей данными в хранилище.

4. Простое уменьшение объема хранимых данных связано с удалением копий документов, расположенных в разных хранилищах. В наилучшем случае ограничение (29) преобразуется в

$$\sum_{i=1}^N x_{ij} = 1 \quad \forall d_j \in H_i. \quad (33)$$

Однако следует помнить о том, что удаление копий может привести к снижению надежности хранения данных. Другая сторона этого шага — необходимость обеспечения проверки корректности (и целостности) данных, помещаемых в хранилище. Эту проблему можно решить выбором надежной (промышленной) СУБД и ее “точной” настройки на решение конкретной задачи. На первый план в этом случае выходит задача администрирования СУБД, а это легче обеспечить при централизованном управлении всей системой электронного документооборота.

5. Стоимость хранения единицы информации в  $i$ -м хранилище  $S_i^{xp}$  зависит от физических характеристик вычислительной системы: стоимости сервера и его обслуживания, стоимости программного обеспечения, а также максимально возможного объема внешней памяти, доступной для хранения данных.

Таким образом, можно с уверенностью сказать, что минимальный вклад этого слагаемого в уравнение (27), принимая во внимание объем документов, с которыми работает пользователь в определенный момент времени, будет обеспечиваться в случае, когда можно отказаться от распределенного хранения документов в силу значимости коэффициента  $S_i^{xp}$  для каждого из хранилищ  $H_i$ . Это приводит к выбору централизованного хранилища, что уменьшает стоимость хранения единицы информации. К тому же такая архитектура системы будет эффективно управляться и, что не менее важно, обновляться с меньшими финансовыми вложениями. Этот фактор может стать определяющим для образовательных учреждений.

**Стоимость получения документов из хранилищ.** Анализируя составляющие для  $S_{пхр}$  в формуле (23), можно выдвинуть следующие предположения.

1. Частота запросов  $\eta_{ij}$  к конкретному документу, определяемому через условие  $x_{ij} \neq 0$  в любой момент времени, ограниченный неравенством (31), будет мала по сравнению с частотой запросов к данному хранилищу  $Q_i$ . Величина  $Q_i$  может быть определена не только задачей документооборота, но и другими прикладными задачами, поскольку хранилища информации используются и другими системами. Кроме того, часть запросов к данным может находиться в данный момент времени в очереди, ожидая своего выполнения (хотя

в число  $Q_i$  эти запросы будут входить). Таким образом, величина  $\frac{1}{Q_i} \sum_{j=1}^{n_i} \eta_{ij} x_{ij} \leq 1$  и не является определяющим фактором для этого слагаемого.

2. Время получения документа из хранилища ( $t_{ij}$ ) определяется двумя факторами: объемом запрошенного документа и пропускной способностью каналов передачи данных. Для рассматриваемой организации пропускная способность каналов передачи данных, как правило, не является сдерживающим фактором, поскольку локальные сети легко модернизируются до требуемых скоростей передачи информации, а “серверный сегмент” такой сети является “выделенным” и связывает необходимые ресурсы серверов информационных хранилищ достаточно мощными каналами. Кроме того, задача решается в условиях существования неравенства (16), которое ограничивает максимальный объем информации, передаваемой пользователем в текущий (достаточно большой) интервал времени  $T$ , определяемый неравенством (31).

Таким образом, в случае устройства распределенного хранилища внутри университета выигрыш по сравнению с выбором единого централизованного хранилища не будет существенным.

В случае же размещения части хранилищ вне сети вуза, например в филиалах, картина будет несколько иной. При таком решении существенный выигрыш в скорости может быть достигнут лишь в случае, когда необходимая информация находится в хранилищах непосредственно этих филиалов. Отсюда следует, что у каждой организации должно быть как минимум одно свое хранилище, в котором будут находиться документы, участвующие в бизнес-процессе этой организации. В противном случае скорость доступа к документам будет резко снижена, поскольку пропускная способность внутренних каналов, как правило, составляет около 100 Мбайт/с, тогда как пропускная способность внешних каналов (между организациями) — около 2 Мбайт/с. В хранилищах вуза имеет смысл содержать только те документы, которые участвуют в его бизнес-процессах или бизнес-процессах, отражающих деятельность вуза и внешних организаций.

Стоимость передачи единицы информации из каждого хранилища определяется двумя составляющими: стоимостью владения коммуникационной сетью  $W_i^{\text{во}}$  на требуемом участке сети и тарифным планом провайдера  $W_i^{\text{т}}$ , у которого арендуется канал передачи данных. Минимизация этих составляющих приводит к целесообразности размещения данных внутри локальной сети организации. При выборе архитектуры системы с централизованным хранилищем вариации стоимости передачи единицы информации зависят только от вариаций стоимости коммуникационного оборудования и расходов на его обслуживание и модернизацию.

Оценивая в совокупности описанные выше предположения, можно сделать вывод, что при централизованном хранении и передаче информации внутри локальной сети организации рассматриваемое слагаемое целевой функции будет давать минимальный вклад в уравнение (27). Следует также отметить, что он будет существенно меньше по сравнению с вкладом первого слагаемого в данное уравнение.

**Стоимость получения документов из “внешних” информационных систем.** Допущения и выводы, сделанные в предыдущем подразделе, могут быть применены для анализа слагаемого целевой функции, которое задано уравнением (24). Разница получения документов из хранилища и из внешней по отношению к СЭД системы незначительна. Однако для более конкретной оценки вклада этого слагаемого необходимо учитывать два обстоятельства.

1. Время получения документа из информационной системы будет неявно включать время поиска (а при необходимости и формирования) запрошенной информации. Кроме того, необходимо учитывать, что запрашивает информацию не только пользователь, но и другие информационные системы. Уменьшить влияние этого обстоятельства можно несколькими путями, например интегрировать хранилище документов с хранилищем “внешней” информационной системы.

2. К стоимости владения каналами передачи данных к “внешней” системе будет добавляться и стоимость владения самой этой системой  $W_i^{по}$ . И этот фактор может стать существенным. Для компенсации его влияния разумным решением станет использование одних и тех же технологий при реализации такой системы или использование свободно распространяемого программного обеспечения.

Как и для предыдущего случая, архитектурное решение с использованием одного централизованного сервера для размещения данных “внешних” систем будет наилучшим, если нет веских аргументов за эксплуатацию и проектирование отдельных внешних систем.

## 2.8. Алгоритм выбора оптимальной архитектуры СЭД

Использование целевой функции (27) для оптимизации архитектуры проектируемой системы документооборота можно осуществлять по описанному ниже алгоритму.

1. На этапе анализа применимости требований к СЭД, а также предметной области необходимо определить оценки сверху для таких величин и свойств, как средний объем электронных документов  $\bar{V}_i$ , которые могут потребоваться для автоматизации бизнес-процессов; максимальная пропускная способность самого медленного участка передачи данных при наибольшей частоте обращения к хранилищу информации изнутри этого сетевого сегмента  $\bar{R}_i$ ; существующие ресурсы вычислительных систем и способности их модернизации для удовлетворения требований пользователей; необходимость сохранения копий документов в распределенном хранилище; необходимость модернизации сети передачи данных и ее стоимости; приемлемое время отклика СЭД на запрос документа  $\bar{T}$ .

2. На основании полученных оценок вычислить: максимальный объем информации  $V^*$ , передаваемой по сети передачи данных, с учетом ограничения (31); максимальный объем хранилищ информации в соответствии с требованиями пользователей и с учетом ограничения (32).

3. На этапе архитектурного проектирования получить оценки для максимальной частоты запросов к хранилищам данных и “внешним” информационным системам.

4. Подсчитать максимальные (оценочные) значения для  $S_{хр}^*$ ,  $S_{пхр}^*$ ,  $S_{пиф}^*$ .

5. Управляя коэффициентами  $\omega_i (i = \overline{1, 3})$ , получить различные вариации значения целевой функции (27) с учетом различных предлагаемых архитектурных решений.

6. Из полученных значений целевой функции  $f^k (k = \overline{1, K})$  выбрать искомое, пользуясь формулой

$$f = \min \{ f^k, k = \overline{1, K} \}. \quad (34)$$

7. Полученное значение целевой функции обеспечит решение задачи оптимизации при любых соотношениях матриц  $X$  и  $Y$ , поскольку решение уравнения (27) мы получили исходя из оценок сверху.

Использование при проектировании СЭД развитой CASE-системы позволит построить диаграммы на основании функциональных моделей и моделей потоков данных, которые могут стать основой для проведения имитационного моделирования с использованием,

например, аппарата сетей Петри, для подтверждения правильности выбора архитектуры будущей системы документооборота.

## 2.9. Выбор оптимальной архитектуры СЭД

Сформулированные требования и определение структуры электронного документа дали первые предпосылки выбора архитектуры СЭД. Для хранения содержимого электронного документа необходима база данных, позволяющая хранить документы различного типа и объема, а также осуществлять контекстный поиск по любым текстовым данным. Поскольку основная работа при поиске нужного электронного документа, какой-либо информации, а также при взаимодействии систем друг с другом будет происходить с метаданными электронных документов, для их хранения необходимо выбрать базу данных, способную обеспечить надежность хранения и обработки большого количества данных, а также приемлемую скорость работы с ними. В качестве такой базы выбрана объектно-реляционная база данных.

В соответствии с приведенным выше алгоритмом произведен выбор архитектуры СЭД. Принято решение о построении СЭД на основе единого централизованного хранилища, содержащего все документы без дублирования. Хранилище является в некотором смысле распределенным и состоит из двух частей — реляционной базы данных, предназначенной для хранения метаданных, и хранилища, ориентированного на работу с документами произвольного типа. Канал связи между этими двумя хранилищами должен обладать максимально возможной пропускной способностью.

## Заключение

Результатом работы является модель оптимизации документопотоков на основе методов математического моделирования, использующая архитектурный подход. Разработан алгоритм выбора оптимальной архитектуры системы электронного документооборота. На основании данного алгоритма произведен выбор архитектуры системы электронного документооборота, разрабатываемой на данный момент в КемГУ.

## Список литературы

- [1] ПРОБЛЕМЫ и типовые решения создания информационной инфраструктуры регионального образовательного комплекса / К.Е. Афанасьев, А.М. Гудов, Ю.А. Захаров и др. Кемерово: Изд-во КемГУ, 2001.
- [2] Гудов А.М., Завозкин С.Ю., Семехина М.В. Об одной модели электронного документооборота вуза // Матер. VIII Междунар. конф. по электронным публикациям "El-Pub 2003". Новосибирск, 2003.
- [3] Гудов А.М., Завозкин С.Ю. Система электронного документооборота // Матер. XXV Всерос. научно-метод. конф. КемГУ "Проблемы обеспечения качества образования". Кемерово, 2004.
- [4] Храпцовская Н.А. Что такое "документ"? // Делопроизводство и документооборот на предприятии. 2004. № 7.

- [5] БАРАХНИН В.Б., ЛЕОНОВА Ю.В. Информационная модель отношений между документами в информационной системе // Вычисл. технологии. 2005. Т. 10, спецвыпуск. С. 129–137.
- [6] БЕССОН С.В. Оптимизация электронного документооборота в корпоративных системах: Дис. ... канд. экон. наук. М., 2001.
- [7] МАТЕМАТИЧЕСКАЯ теория оптимальных процессов / Л.С. Понтрягин, В.Г. Болтянский, Р.В. Гамкрелидзе, Е.Ф. Мищенко. М.: Наука, 1983.
- [8] МОИСЕЕВ Н.Н., ИВАНИЛОВ Ю.П., СТОЛЯРОВА Е.М. Методы оптимизации. М.: Наука, 1983.

*Поступила в редакцию 20 февраля 2006 г.,  
в переработанном виде — 16 марта 2006 г.*