

УДК 651.4/9:004

МОДЕЛЬ РЕВЕРСИВНОГО ДВИЖЕНИЯ ДОКУМЕНТОВ В СИСТЕМЕ УПРАВЛЕНИЯ ДОКУМЕНТАМИ DIRECTUM

Ф. Ф. Иванов, К. Г. Семенов

*Сургутский государственный университет
iff777@yandex.ru, SemenovK@mail.ru*

В данной статье рассмотрена организация хранения текстов электронных документов в системе управления документами DIRECTUM, а также модель реверсивного перемещения документов. Описано условие размещения текстов электронных документов в оперативных и архивных хранилищах. Предложен подход к рациональному распределению текстов документов между ними. Представленный подход может быть реализован с применением программного робота, что позволит не затрагивать системную функциональность платформы системы управления документами DIRECTUM.

Ключевые слова: архивы электронных документов, прогнозирование, математические модели, модели потоков документов, DIRECTUM, программный робот.

MODEL OF REVERSE MOVEMENT OF DOCUMENTS IN THE DOCUMENT MANAGEMENT SYSTEM DIRECTUM

F. F. Ivanov, K. G. Semenov

*Surgut State University
iff777@yandex.com, SemenovK@mail.ru*

The article considers the organization of the storage of electronic documents in the document management system DIRECTUM and the model of the reverse movement of documents as well. The condition for placing texts of electronic documents in operational and archival repositories is described. An approach to the smart distribution of the texts of documents between them is proposed. The proposed approach can be implemented using a software robot, which will not affect the system functionality of the document management system platform DIRECTUM.

Keywords: archives of electronic documents, forecasting, mathematical models, document flows models, DIRECTUM, software robot.

Каждый документ в полнотекстовой системе DIRECTUM имеет карточку и текст. В карточке хранятся значения выделенных основных реквизитов, таких как наименование, автор, приложение и прочее. Текст – это текст документа в формате приложения-редактора.

Карточки всех документов системы DIRECTUM хранятся в базе данных SQL-сервера. Тексты документов размещаются в хранилищах текстов документов – файловых хранилищах документов [1, 2]. На рис. 1 изображена схема и описание назначения хранилищ текстов документов.

Основным хранилищем текстов является оперативное файловое хранилище документов. Это хранилище всегда есть в системе DIRECTUM. В этом хранилище по умолчанию размещаются тексты всех документов.

Файловые хранилища предназначены для размещения текстов документов в файловых системах одного или нескольких компьютеров. С одной стороны, это позволяет уменьшить объем БД SQL-сервера, снизив тем самым требования к используемым аппаратным средствам, и, как следствие, уменьшить совокупную стоимость владения системой DIRECTUM, с другой – предоставляет широкие возможности в организации работы с документами:

- в файловых хранилищах можно размещать документы больших объемов, например, фото- и видеоматериалы, размер документа в таких хранилищах ограничен только размером дискового пространства;
- можно настроить автоматический перенос текстов документов в архивные хранилища и обратно согласно заданным условиям.

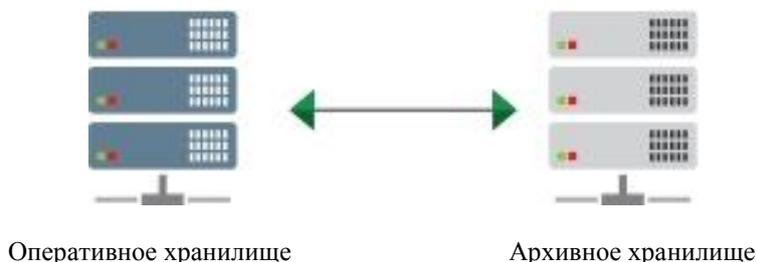


Рис. 1. Файловые хранилища СУД DIRECTUM

Тексты документов из оперативных хранилищ, в которых они размещаются сразу после создания документа, могут автоматически переноситься в архивные хранилища. Условия переноса документов задаются в справочнике «Правила архивирования». Сценарий «Агент архивирования» переносит документы в архивное хранилище.

Архивное хранилище, в которое будут перемещаться тексты документов, задается в записи справочника «Правила архивирования».

В архивных хранилищах тексты документов доступны только для просмотра. Если возникает потребность изменить документ, например, отредактировать текст или создать версию документа, то для выполнения этих действий пользователю нужно подтвердить операцию извлечения документа из архивного хранилища [1].

Основными требованиями к хранилищам документов относительно их типов (оперативное и архивное) являются следующие:

- для оперативного хранилища основным требованием является обеспечение необходимой нагрузочной способности этого хранилища при работе системы с текстами электронных документов (запись, чтение, модификация);
- для архивного хранилища основным требованием является обеспечение минимальной стоимости хранения текстов документов.

Поэтому критерием размещения текста электронного документа в архивном хранилище, в первую очередь, является низкая востребованность этого документа пользователями системы электронного документооборота и непосредственно самой системой документооборота, так как она в автоматическом режиме может производить различные операции над текстами документов.

Оценить востребованность текста электронного документа в СУД DIRECTUM возможно посредством анализа данных таблицы **SBEDocProtocol**, которая содержит протокол работы с документами. Данная таблица обновляется при любых действиях с документами. В таблице представлен набор полей, достаточный для оценки востребованности электронного документа.

Таблица

Состав полей таблицы протокола работы с документами

Поле	Описание
EDocID	ИД документа
Version	Версия документа
UserID	ИД пользователя
...	

Окончание табл.

Поле	Описание
ActionType	Действие: «7» – Разрешение передачи прав доступа «8» – Перешифрование «А» – Изменение прав доступа «В» – Изменение видимости версии «С» – Создание «D» – Удаление «Е» – Изменение «F» – Шифрование сертификатом «G» – Подписание «H» – Удаление версии
ActionDate	Дата и время действия
Detail	Детализация данных о действии
XRecID	ИД записи в таблице
...	

В качестве оценки востребованности документа принимается среднее количество обращений к одному документу за период в одну неделю. Период в одну неделю выбран из соображений необходимой частоты перемещения текста документа из оперативного хранилища в архивное.

$$K_{\text{востреб}}(T_{\text{возвр}}) = \frac{K_{\text{востреб всех}}(T_{\text{возр}})}{K_{\text{док}}} \quad (1)$$

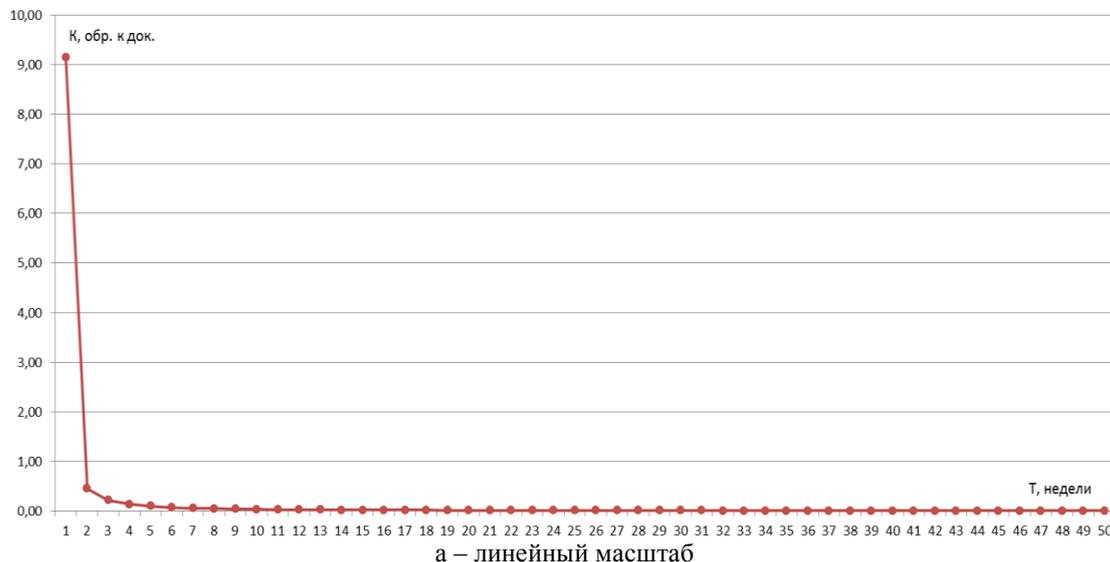
где $T_{\text{возр}}$ – время после создания документа («возраст документа»);

$K_{\text{востреб}}(T_{\text{возр}})$ – среднее количество обращений к одному документу за контрольный период (1 неделя);

$K_{\text{востреб всех}}(T_{\text{возр}})$ – среднее количество обращений ко всем документам за контрольный период (1 неделя);

$K_{\text{док}}$ – количество документов.

Критерий $K_{\text{востреб}}(T_{\text{возр}})$ зависит от времени $T_{\text{возр}}$, прошедшего с момента создания документа, и по сути является возрастом документа. С целью выявления зависимости критерия от времени существования документов проводится разведочный анализ по всем документам независимо от типа. Его результаты представляются в виде графика.



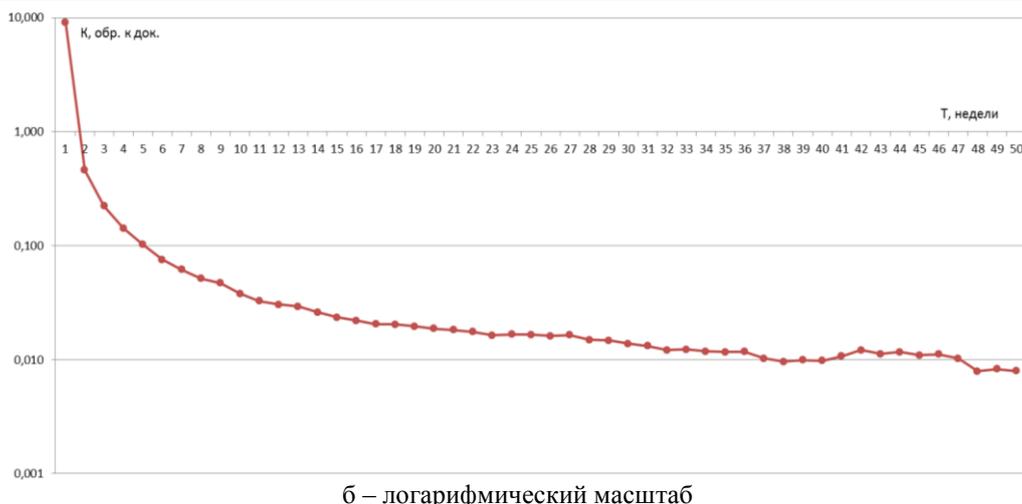


Рис. 2. Среднее количество обращений к документу в течение одной недели от «возраста документа»: а – линейный масштаб; б – логарифмический масштаб

На рис. 2 отображено среднее количество обращений к документу за одну неделю возраста документа, где видно, что кривая зависимости критерия $K_{\text{востреб}}$ от времени достаточно гладкая и монотонная.

Данная зависимость может быть хорошо отражена трендовой моделью.

Трендовая модель – это математическая модель, описывающая изменение прогнозируемого или анализируемого показателя только в зависимости от времени и имеющая вид: $y = f(t)$.

Наиболее распространенным методом прогнозирования является нахождение аналитического выражения (уравнения) тренда [3–5]. Тренд экстраполируемого процесса либо явления – это основная тенденция временного ряда, в некоторой мере свободная от случайных воздействий.

Разработка прогноза заключается в определении вида экстраполирующей функции $y = f(t)$, которая выражает зависимость изучаемой величины от времени на основе исходных наблюдаемых данных. Первым этапом является выбор оптимального вида функции, дающей наилучшее описание тренда. Наиболее часто используются следующие зависимости [3–9]:

- линейная: $y = b_0 + b_1t$; (2)

- параболическая: $y = b_0 + b_1t + b_2t^2$; (3)

- показательная функция: $y = b_0b_1^t$. (4)

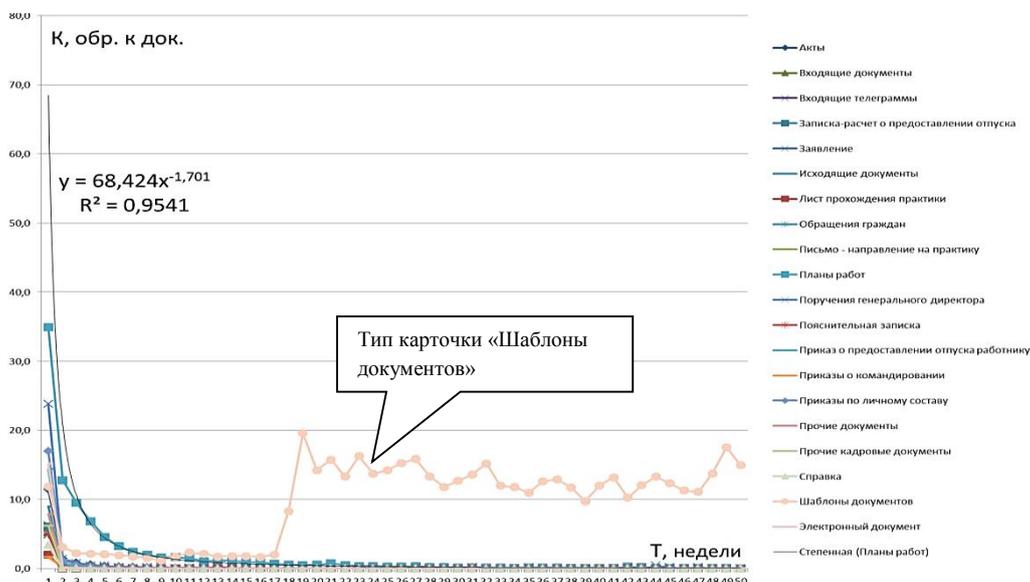


Рис. 3. Среднее количество обращений к документу в течение одной недели от «возраста документа» в разрезе типов карточек документов

На рис. 3 показано среднее количество обращений к документу за одну неделю от возраста документа в разрезе типов документов. Видно, что некоторые кривые не отражают общую тенденцию убывания количества обращений во времени, например, документы с типом карточки «Шаблоны документов» выбиваются из общей тенденции поведения документов.

Известны причины этого. Кривая, отражающая обращения к документам с типом карточки «Шаблоны документов», имеет до 17-й недели сравнительно небольшие значения: 2–3 обращения в неделю, а далее кратно увеличивается и стабилизируется на уровне 15 обращений в неделю. Это объясняется тем, что этот тип документа используется именно по своему названию – как шаблон для создания текста нового документа, и поэтому количество обращений к нему в общем случае будет постоянным. Некоторая задержка в востребованности документа объясняется тем, что новый шаблон некоторое время «приживается» в бизнес-процессе предприятия.

Пример поведения количества обращений к электронному документу с типом карточки «Шаблоны документов» показывает, что в целом тренд критерия будет определяться «бизнес-сущностью» документа и зависеть от значения определенных реквизитов карточки документа.

Основными реквизитами карточки документа, определяющими прикладное назначение документа, являются:

- тип карточки документа;
- вид документа.

Кроме того, такими реквизитами могут являться «стадия жизненного цикла», «автор документа (его структурное подразделение)», «приложение редактор» и другие прикладные реквизиты. Влияние на поведение тренда может оказывать наличие ссылок на документ в других объектах СУД DIRECTUM. Влияние этих факторов на тренд критерия необходимо предварительно определять.

На рис. 4 показан процесс подготовки (вычисления) параметров системы для осуществления переноса в архив и возврата документов из архива. Непосредственно сам перенос документов в архивное хранилище является функцией системы управления документами DIRECTUM. Для его осуществления должны быть определены параметры архивирования.

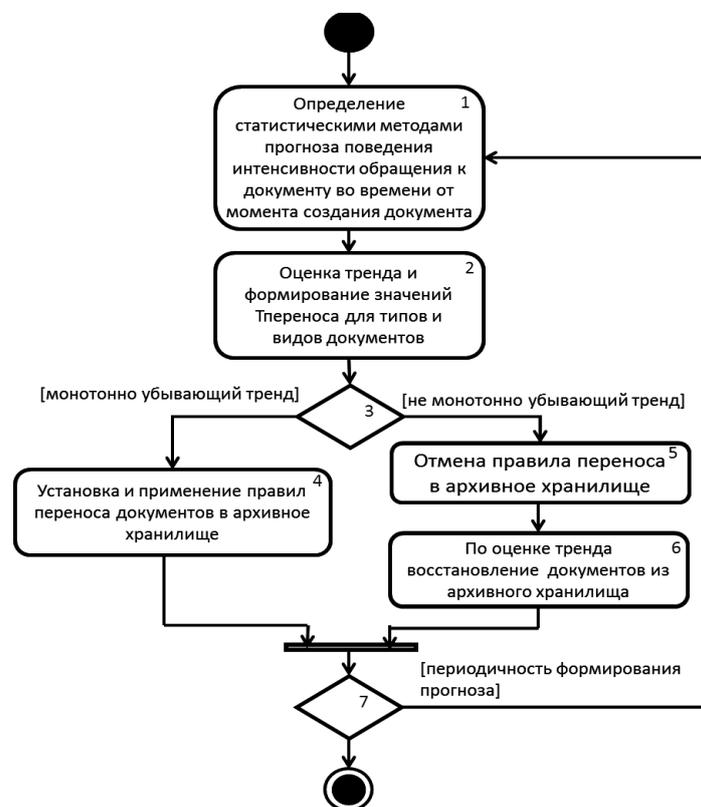


Рис. 4. Процесс-модель подготовки (вычисления) параметров системы для осуществления переноса в архив и возврата документов из архива (реверса)

Алгоритм более детально может быть описан таким образом:

- блок 1 – осуществляет процедуру аппроксимации востребованности к документу для различных видов и типов документов;
- блок 2 – по полученным аналитическим функциям определяется целесообразность осуществления архивирования документа данного вида и, (в случае целесообразности) определяется Т-переноса – параметр для определения условий переноса в архивное хранилище, то есть документы определенного типа будут перенесены в архивное хранилище, если их возраст (время прошедшее после создания) превысило этот параметр;
- блок 3 – осуществляет определение направления движения документов в суд *directum* для видов и типов электронных документов;
- блок 4 – осуществляет установку параметров автоматического архивирования определенных типов и видов документов;
- блок 5 – производит отмену правил архивирования определенных типов и видов электронных документов;
- блок 6 – осуществляет оценку тренда критерия востребованности данного типа, вида документов и осуществляет восстановление их из архивного хранилища;
- блок 7 – осуществляет периодическое повторение этих процедур с целью отслеживания актуальности примененных параметров реверсивного движения документов.

Заключение. Критерием необходимости перемещения текста электронного документа в архивное хранилище можно принять среднее количество обращений к одному документу за определенный (контрольный) период. Используя статистическую адаптивную (трендовую) модель прогнозируемого значения этого критерия от времени после создания документа (возраста документа), возможно прогнозировать снижение интенсивности обращения к электронному документу и устанавливать порог «возраста документа» (Т - переноса), после которого возможно осуществить перенос документа из оперативного хранилища в архивное хранилище, не снижая эксплуатационных характеристик системы управления документами.

Данный подход позволит обеспечить:

- 1) управляемый размер оперативного хранилища;
- 2) в целом обеспечит снижение стоимости хранения электронных документов за счет размещения мало востребованных документов в хранилище с низкой стоимостью хранения;
- 3) по оценке статистической модели в различных разрезах (по видам, типам электронных документов), возможно осуществлять возврат (реверс) документов из архивного хранилища в оперативное для обеспечения возможности интенсивной работы с документами;
- 4) данная модель реверсивного движения документов будет использована в качестве функциональности программного робота СУД *DIRECTUM*.

В дальнейшем предполагается развить математические модели и механизм вычисления параметров модели и технологию перемещения документов с оценкой экономических показателей функционирования в составе СУД *DIRECTUM*.

Литература

1. *Directum* – корпоративная система электронного документооборота. URL: <http://www.directum.ru/315540.aspx> (дата обращения: 12.10.2018).
2. Перспективы Workflow в России // Еженедельник «ComputerworldРоссия». 2000. № 13. URL: <https://www.osp.ru/cw/2000/13/4101/> (дата обращения: 10.10.2018).
3. Койчубеков Б. К. Биостатистика. Алматы : КГМУ, 2014. 80. 47 с.
4. Лебедева И. М., Федорова А. Ю. Макроэкономическое планирование и прогнозирование. СПб. : Ун-т ИТМО, 2016. С. 23–27.
5. Снитюк В. Е. Прогнозирование. Модели, Методы, Алгоритмы. Киев, 2008. С. 125.
6. Ахмад Б., Кучуганов В. Н. Некоторые подходы к интеллектуализации систем офисного документооборота // Перспективные информационные технологии и интеллектуальные системы. 2003. № 4 (16). С. 72–81.

7. Гмурман В. Е. Теория вероятностей и математическая статистика. 9-е изд., перераб. и доп. М. : Высш. школа, 2003. 479 с.
8. Роберте Ф. С. Дискретные математические модели с приложениями к социальным, биологическим и экономическим задачам. М. : Наука, 1986. 496 с.
9. Поллард Дж. Справочник по вычислительным методам статистики. М. : Финансы и статистика, 1982. 344 с.