УДК 621.377

В.В. Типикин, М.А. Скибина

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ В СИСТЕМЕ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА

Типикин Всеволод Валентинович, кандидат технических наук, окончил физико-технический факультет Ульяновского государственного университета. Заместитель главного конструктора ФНПЦ АО «НПО «Марс». Имеет статьи в области автоматизации проектирования информационных систем. [e-mail: mars@mv.ru].

Скибина Мария Анатольевна, кандидат физико-математических наук, окончила механико-математический факультет УлГУ. Ведущий инженер ФНПЦ АО «НПО «Марс». Имеет статьи в области исследования применимости генетических алгоритмов к проблеме оптимизации стохастических систем. [e-mail: mars@mv.ru].

Аннотация

В статье выполнен обзор систем электронного документооборота, выявлена перспективность использования интеллектуальных инструментов. Рассмотрена система электронного документооборота, разработанная ФНПЦ АО «НПО «Марс» и используемая в изделиях предприятия. Предлагается расширить функциональные возможности системы за счет включения подсистемы интеллектуального анализа данных. Предполагается собирать данные в процессе эксплуатации системы и выполнять их анализ. Определен набор данных, характеризующих процесс разработки документов: время разработки документов плановое и фактическое; объем разработанных документов; качество разработанных документов. Под анализом понимается установление корреляционных и регрессионных зависимостей между характеристиками. Результаты анализа должны использоваться должностными лицами органов управления для оптимальной организации выполнения работ по подготовке документов. Приведен пример установления регрессионных зависимостей между временем разработки одного типа документа и его объемом. Случайные значения показателей установлены путем имитационного моделирования.

Ключевые слова: электронный документооборот, интеллектуальный анализ данных, регрессионные зависимости.

doi: 10.35752/1991-2927-2020-1-5-65-74

DATA MINING IN DOCUMENT AUTOMATION SYSTEM

Vsevolod Valentinovich Tipikin, Candidate of Sciences in Engineering; graduated from the Faculty of Physics and Mathematics of Ulyanovsk State University; Deputy Chief Designer at FRPC JSC 'RPA 'Mars'; an author of articles in the field of automation of information system design. e-mail: mars@mv.ru.

Maria Anatolevna Skibina, Candidate of Sciences in Physics and Mathematics; graduated from the Faculty of Mechanics and Mathematics of Ulyanovsk State University; Leading Engineer of FRPC JSC 'RPA 'Mars'; an author of articles in the field of genetic algorithms applicability to stochastic systems optimization. e-mail: mars@mv.ru.

Abstract

The paper reviews the document automation systems and determines the perspectives of intellectual tools usage. It considers the document automation system developed by FRPC JSC 'RPA 'Mars' and implemented in its products. The system functionality is offered to be improved by adding the data mining subsystem. The subsystem intends to accumulate data while in service and to analyze them. The paper specifies the data set, defining the process of document developing, i.e. planned and actual time of developing, amount and quality of finished documents. Data mining implies a determination of correlation and regression dependencies between attributes. Based on the results of data mining, management officials are able to arrange the process of document developing effectively. The authors give an example of regression dependencies between the time taken to develop a document of one type and its size. Random values of performances are given while simulation modeling.

Key words: document automation, data mining, regression dependencies.

ВВЕДЕНИЕ

Система электронного документооборота (СЭД) – автоматизированная многопользовательская система, сопровождающая процесс управления работой иерархической организации с целью обеспечения выполнения этой организацией своих функций [1]. СЭД является одной из важных составляющих ECM-системы (Enterprise Content Management), основная суть которой заключается в управлении всевозможным корпоративным контентом и бизнес-процессами предприятия.

В ГОСТ Р 7.0.8-2013 документооборот определен как движение документов в организации с момента их создания или получения до завершения исполнения или отправления и представляет собой комплекс работ с документами: прием, регистрация, рассылка, контроль исполнения, формирование дел, хранение и повторное использование документации, справочная работа.

Основными принципами, которыми руководствуются при создании СЭД, являются [2, 3]:

- однократная регистрация документа, позволяющая однозначно идентифицировать документ;
- возможность параллельного выполнения операций, позволяющая сократить время движения документов и повысить оперативность их исполнения;
- непрерывность движения документа, позволяющая идентифицировать ответственного за исполнение документа (задачи) в каждый момент времени жизни документа (процесса);
- единая (или согласованная распределенная) база документной информации, позволяющая исключить возможность дублирования документов;
- эффективно организованная система поиска документа, позволяющая находить документ, обладая минимальной информацией о нем;
- развитая система отчетности по различным статусам и атрибутам документов, позволяющая контролировать движение документов по процессам документооборота и принимать управленческие решения, основываясь на данных из отчетов.

В статье рассматривается расширение функциональных возможностей СЭД за счет включения подсистемы интеллектуального анализа данных. Данными являются параметры процесса разработки документов, устанавливаемые в ходе эксплуатации. Интеллектуальный анализ предполагает установление зависимостей между параметрами для формальной постановки и решения задач оптимальной организации работ.

1 ОПИСАНИЕ СЭД

Российский рынок СЭД и услуг по их внедрению из года в год растет [4] (рис. 1). Росту рынка способствует не только движение в сфере импортозамещения, но и реальная потребность бизнеса в более функциональных и масштабируемых платформах, на которых возможно оперативно разработать системы управления различными бизнес-процессами и создать уникальные решения под специфические требования заказчиков.

По количеству реализованных проектов лидируют СЭД Directum, Elma, Docsvision, «Дело» и «Тезис».

В СЭД, представленных на отечественном рынке, уже делаются попытки использовать интеллектуальный анализ данных для ускорения процесса обработки документа [5]. Компания «Диджитал Дизайн» применила интеллектуальные алгоритмы для кластеризации документов, предсказания атрибутов документов, автоматического реферирования и выделения аномалий в документах в системе документационного управления (СДУ) «Приоритет» на платформе DocsVision. Это позволит системе самостоятельно определить тип документа и организовать процесс разработки, согласования и утверждения с минимальным участием человека.

Продукты компании TerraLink активно используют технологии машинного обучения в задачах распознавания входящих документов, в дальнейшем компанией планируется применение таких технологий для оценки эффективности процесса и поиска новых идей по оптимизации [6].

Разработанное в 2018 году решение для интеллектуальной обработки корпоративного контента Directum Ario включает набор интеллектуальных инструментов, использующих алгоритмы машинного обучения, который предназначен для выполнения рутинных операций: распределения потока документов, заполнения реквизитов, регистрации и отправки исполнителям [7]. По словам разработчиков, сервис классификации обучается на документах, которые уже есть в организации, и, таким образом, учитывает особенности делопроизводства. Пользователям остается лишь контролировать результат выполненной работы.

В ПАО «Интелтех» коллективом авторов СЭД «Цера» была предпринята попытка разработки инфраструктуры глубинного анализа процессов [8]. Реализованный модуль построения модели управления потоком документов осуществляет процесс сегментирования журнала событий на основе выбранного типа документа и стадии его жизненного цикла и производит реконструкцию графа рабочего процесса нотации BPMN (Business process management notation). В дальнейшем авторы

Рис. 1. Динамика российского рынка СЭД, млрд. руб.

предполагают использовать алгоритм индуктивного анализа.

ФНПЦ АО «НПО «Марс» была разработана СЭД (рис. 2), активно используемая на самом предприятии и на объектах заказчика (около 2000 пользователей), предоставляющая следующие возможности [9]:

- создание документа путем присоединения вложенных файлов, разработанных ранее с помощью пакета офисных средств, и ввод его в электронный документооборот (ЭДО);
- объединение в одном документе информационных материалов (документов, презентаций, клипов, рисунков, заданий и т. д.) различного характера;
- ввод в ЭДО документа, поступившего из каналов связи;
- ввод в ЭДО документа, разработанного неавтоматизированным способом (на бумаге), путем сканирования текста и заполнения вручную результатов его обработки;
- обработка документа (согласование, утверждение, исполнение резолюций);

- оповещение пользователя о результатах обработки документов, направленных им на согласование, утверждение, исполнение резолюций в адрес других должностных лиц (ДЛ);
- автоматический контроль сроков исполнения резолюций руководства и оповещение пользователей о документах с критическими сроками исполнения;
- поиск документов в базе данных, возможность сохранения наиболее часто используемых условий отбора;
- выдача документа в каналы связи;
- публикация документа на специальном сайте;
- передача отработанного документа в электронный архив;
- подключение приложений для обработки вложений в зависимости от их типа (расширения): офисные средства, графические пакеты либо специальные приложения пользователей, в том числе обновление баз данных с помощью вложений документов;
- настройка служебной информации, необходимой для функционирования СЭД, на конкретный объект автоматизации, в том числе определение структуры

Рис. 2. Процесс работы с документами

организации, описание пользователей и их полномочий, описание внешних систем, а также ведение словарей документооборота;

- настройка внешнего вида СЭД.

В основе СЭД лежит автоматизация основных процедур, выполняющих необходимое взаимодействие между исполнителями и утверждающими ДЛ. Обмен всей информацией осуществляется по сети, без непосредственной встречи участников процесса. В дальнейшем утвержденный документ может быть распечатан и подписан, а также сохранен в бумажном виде в архиве.

Приведенное описание является упрощенной схемой применения СЭД. Дополнительные функциональные возможности предоставляются такими средствами, как обмен оповещениями по сети, электронная почта, электронный архив документов, регламентированный документооборот.

Последнее из перечисленных средств представляет собой надстройку к СЭД, предназначенную для описания совокупности правил, определяющих порядок формирования набора связанных документов (регламента), и содержит следующие этапы: планирование, выполнение и контроль разработки.

На этапе планирования пользователь в роли администратора, имеющий полный доступ, с помощью средств и инструментов регламентированного документооборота разрабатывает план-график создания документов, создает этапы, добавляет к ним документы, назначает ответственных исполнителей, согласующих, утверждающих, определяет время, необходимое для разработки документа. За основу нового плана-графика может быть взят существующий с сохранением ранее разработанных документов в виде шаблонов и с сохранением назначенного доступа пользователей. Разработанный план-график отображается в виде дерева, а также в виде диаграммы Ганта, при этом пользователю предоставляется возможность настроить вид отображения плана-графика в соответствии со своими предпочтениями.

Выполнение плана-графика начинается с момента его запуска администратором или другим пользователем, имеющим соответствующие права. В процессе запуска план-график проходит проверку на корректность по определенным критериям. Если план-график содержит ошибки, пользователь получает сообщение об этом и ему выдается список всех некорректных элементов плана-графика с подробным описанием ошибок. После успешного запуска плана-графика ответственные исполнители получают соответствующее оповещение и могут приступить к оформлению, отправке на согласование и утверждение документов средствами ЭДО.

Контроль разработки плана-графика реализован следующим образом. Все пользователи, имеющие доступ хотя бы к одному документу, могут наблюдать ход выполнения плана-графика в системе регламентированного документооборота. Пользователю предоставляется возможность увидеть статистику выполнения плана-графика в целом, также можно просмотреть статистику разработки документов по конкретному пользователю.

В качестве возможных вариантов расширения функциональности СЭД был предложен алгоритм оптимального распределения ресурсов методами имитационного моделирования и случайного поиска [10, 11]. В [12, 13] предлагается использовать СЭД для рациональной реализации функциональных систем на основе построенных математических зависимостей между характеристиками процессов разработки документов.

В статье рассматривается возможность развития ЭДО за счет включения подсистемы интеллектуального анализа данных (рис. 3).

Рис. 3. Предлагаемая схема работы с регламентированным документооборотом

Интеллектуальный анализ данных заключается в поиске функциональных и логических закономерностей в накопленных данных, построении моделей и правил, которые объясняют найденные закономерности и/или (с определенной вероятностью) прогнозируют развитие некоторых процессов. Применение интеллектуального анализа данных в СЭД позволит определить типы документов и этапы разработки документов, наиболее требовательные к ресурсам, необходимым для выполнения, а также найти способы для ускорения разработки документов и повышения их качества.

2 СБОР ДАННЫХ ДОКУМЕНТООБОРОТА

Подсистема интеллектуального анализа данных должна быть частью СЭД и выполнять следующие функции:

- сбор данных о разработке документов;
- установление вероятностных характеристик полученных данных;
- установление корреляционных и регрессионных зависимостей между различными показателями разработки документов;
- представление результатов анализа пользователю для формирования решений по оптимальной организации работ.

База администрирования информационных ресурсов

База документов

Рис. 4. Базы данных СЭД

СЭД постоянно ведется сохранение в базе документов данных, характеризующих процесс разработки, согласования и утверждения различных типов документов (рис. 4). Для интеллектуального анализа необходимо расширить содержание базы, включая в состав количественные показатели разработки документов за определенный период времени.

Таковыми данными являются, например, запланированное время разработки документа T_p и фактически затраченное T_f , так как в процессе планирования независимо от навыков пользователя не всегда удастся заранее определить время, необходимое для разработки документа. В то же время важным параметром, показывающим насколько запланированное время отличается от фактического, является разность $T_d = T_p - T_f$.

Для анализа необходимо иметь следующие выборки данных:

$$\begin{aligned} T_p &= \{T_{p1}, T_{p2}, \dots, T_{pn}\}; \\ T_f &= \{T_{f1}, T_{f2}, \dots, T_{fn}\}; \\ T_{d1} &= T_{p1} - T_{f1}, T_{d2} = T_{p2} - T_{f2}, \dots, T_{dn} = \\ &= T_{pn} - T_{fn}; \\ T_d &= \{T_{d1}, T_{d2}, \dots, T_{dn}\}, \end{aligned}$$

где n – количество данных в выборке.

Все документы могут быть разбиты на группы (например, по расширению файла), что позволит выполнить анализ отдельно для каждой группы:

$$\begin{aligned} T_{pi} &= \{T_{pi1}, T_{pi2}, \dots, T_{pin}\}; \\ T_{fi} &= \{T_{fi1}, T_{fi2}, \dots, T_{fin}\}; \\ T_{di1} &= T_{pi1} - T_{fi1}, T_{di2} = T_{pi2} - T_{fi2}, \dots, T_{din} = \\ &= T_{pin} - T_{fin}; \\ T_{di} &= \{T_{di1}, T_{di2}, \dots, T_{din}\}; \\ i &= [1, m], \end{aligned}$$

где i – номер типа документа, m – количество типов документов.

С учетом этапности выполнения документов выборки будут иметь вид:

$$\begin{aligned} T_{pij} &= \{T_{pij1}, T_{pij2}, \dots, T_{pijn}\}; \\ T_{fij} &= \{T_{fij1}, T_{fij2}, \dots, T_{fijn}\}; \\ T_{dij1} &= T_{pij1} - T_{fij1}, T_{dij2} = T_{pij2} - T_{fij2}, \dots, T_{dijn} = \\ &= T_{pijn} - T_{fijn}; \\ T_{dij} &= \{T_{dij1}, T_{dij2}, \dots, T_{dijn}\}; \\ j &= [1, k], \end{aligned}$$

где j – номер этапа разработки документа, k – количество этапов.

Далее введем понятие набора документов, для него также можно определить планируемое время выполнения TN_{pi} и фактическое TN_{fi} , разницу между ними TN_{di} . Тогда соответствующие выборки данных для наборов имеют следующий вид:

$$\begin{aligned}
TN_{pi} &= \{TN_{pi1}, TN_{pi2}, \dots, TN_{pin}\}; \\
TN_{fi} &= \{TN_{fi1}, TN_{fi2}, \dots, TN_{fin}\}; \\
TN_{di1} &= TN_{pi1} - TN_{fi1}, TN_{di2} = TN_{pi2} - TN_{fi2}, \dots, \\
&\dots, TN_{din} = TN_{pin} - TN_{fin}; \\
TN_{di} &= \{TN_{di1}, TN_{di2}, \dots, TN_{din}\}; \\
i &= [1, M],
\end{aligned}$$

где i – номер набора, M – количество наборов.

С учетом этапов выполнения наборов документов необходимо ввести номер этапа:

$$\begin{aligned}
TN_{pij} &= \{TN_{pij1}, TN_{pij2}, \dots, TN_{pijn}\}; \\
TN_{fij} &= \{TN_{fij1}, TN_{fij2}, \dots, TN_{fijn}\}; \\
TN_{dij1} &= TN_{pij1} - TN_{fij1}, \\
TN_{dij2} &= TN_{pij2} - TN_{fij2}, \dots, \\
&\dots, TN_{dijn} = TN_{pijn} - TN_{fijn}; \\
TN_{dij} &= \{TN_{dij1}, TN_{dij2}, \dots, TN_{dijn}\}.
\end{aligned}$$

Связь между параметрами набора документов и параметрами документов, входящих в соответствующий набор:

$$\begin{aligned}
TN_{pi} &= \sum_{j \in J_i} T_{pj}, \quad TN_{fi} = \sum_{j \in J_i} T_{fj}, \\
TN_{di} &= \sum_{j \in J_i} T_{dj},
\end{aligned}$$

где J_i – подмножество номеров документов, входящих в набор.

С оценкой времени разработки связана другая, не менее важная характеристика, отвечающая за объем документа S . Необходимо отметить, что на этапе планирования объем документа может быть оценен лишь приблизительно и на его величину могут повлиять различные факторы. Поэтому при анализе будем учитывать объем только готовых (утвержденных) документов:

$$S = \{S_1, S_2, \dots, S_n\}.$$

В соответствии с разбиением документов по типам выборка будет иметь вид:

$$\begin{aligned}
S_i &= \{S_{i1}, S_{i2}, \dots, S_{in}\}; \\
i &= [1, m].
\end{aligned}$$

Для этапов выполнения документов получаем:

$$\begin{aligned}
S_{ij} &= \{S_{ij1}, S_{ij2}, \dots, S_{ijn}\}; \\
j &= [1, k].
\end{aligned}$$

Объем набора документов обозначим SN_i , тогда выборка будет записана следующим образом:

$$SN_i = \{SN_{i1}, SN_{i2}, \dots, SN_{in}\}.$$

Этапы выполнения наборов документов будут учтены в выборке:

$$SN_{ij} = \{SN_{ij1}, SN_{ij2}, \dots, SN_{ijn}\};$$

$$j = [1, k].$$

Объем конкретного набора документов является суммой объемов входящих в набор документов:

$$SN_i = \sum_{j \in J_i} S_j.$$

В качестве параметра оценки выполнения заданного регламента можно использовать качество разработанного документа Q . Для этого можно воспользоваться услугами эксперта, лица, инициировавшего процесс разработки плана-графика, утверждающего конкретный документ, или любого другого, обладающего достаточным опытом. В таком случае можно рассмотреть выборку:

$$Q = \{Q_1, Q_2, \dots, Q_n\}.$$

Качество типов документов характеризует следующая выборка:

$$\begin{aligned}
Q_i &= \{Q_{i1}, Q_{i2}, \dots, Q_{in}\}; \\
i &= [1, m].
\end{aligned}$$

При анализе этапов выполнения документов необходимо оперировать следующими данными:

$$\begin{aligned}
Q_{ij} &= \{Q_{ij1}, Q_{ij2}, \dots, Q_{ijn}\}; \\
j &= [1, k].
\end{aligned}$$

Для анализа качества набора документов QN_i должна использоваться выборка:

$$\begin{aligned}
QN_i &= \{QN_{i1}, QN_{i2}, \dots, QN_{in}\}; \\
i &= [1, M],
\end{aligned}$$

где i – номер набора, M – количество наборов, n – количество данных в выборке.

Этапы выполнения наборов документов будут учтены в выборке:

$$\begin{aligned}
QN_{ij} &= \{QN_{ij1}, QN_{ij2}, \dots, QN_{ijn}\}; \\
j &= [1, k].
\end{aligned}$$

Объем конкретного набора документов является суммой объемов входящих в набор документов:

$$QN_i = \sum_{j \in J_i} Q_j.$$

3 ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

Интеллектуальный анализ данных разбивается на следующие этапы [14]:

- понимание и формулировка задачи анализа;
- подготовка данных для автоматизированного анализа;
- применение методов Data Mining и построение моделей;
- проверка построенных моделей;
- интерпретация моделей.

Таким образом, имея подготовленные выборки данных СЭД, можно приступить к построению модели, описывающей процесс разработки документов.

Математическое ожидание (среднее значение) находится по следующим формулам [15]:

$$M[T_{pi}] = \frac{1}{n} \sum_{j=1}^n T_{pij}, \quad M[T_{fi}] = \frac{1}{n} \sum_{j=1}^n T_{fij},$$

$$M[T_{di}] = \frac{1}{n} \sum_{j=1}^n T_{dij},$$

$$M[S_i] = \frac{1}{n} \sum_{j=1}^n S_{ij},$$

$$M[Q_i] = \frac{1}{n} \sum_{j=1}^n Q_{ij}.$$

Далее необходимо рассчитать дисперсию и среднее квадратическое отклонение:

$$D[T_{pi}] = \frac{1}{n} \sum_{j=1}^n (T_{pij} - M[T_{pi}])^2,$$

$$\sigma_{T_{pi}} = \sqrt{D[T_{pi}]};$$

$$D[T_{fi}] = \frac{1}{n} \sum_{j=1}^n (T_{fij} - M[T_{fi}])^2,$$

$$\sigma_{T_{fi}} = \sqrt{D[T_{fi}]};$$

$$D[T_{di}] = \frac{1}{n} \sum_{j=1}^n (T_{dij} - M[T_{di}])^2,$$

$$\sigma_{T_{di}} = \sqrt{D[T_{di}]};$$

$$D[S_i] = \frac{1}{n} \sum_{j=1}^n (S_{ij} - M[S_i])^2,$$

$$\sigma_{S_i} = \sqrt{D[S_i]};$$

$$D[Q_i] = \frac{1}{n} \sum_{j=1}^n (Q_{ij} - M[Q_i])^2,$$

$$\sigma_{Q_i} = \sqrt{D[Q_i]}.$$

Функции распределения будут иметь вид:

$$F(T_{pi}) = \begin{cases} 0, & T_{pi} \leq T_{pi(1)}, \\ \frac{j}{n}, & T_{pi(j)} < T_{pi} \leq T_{pi(j+1)}, \\ 1, & T_{pi} > T_{pi(n)}; \end{cases}$$

$$F(T_{fi}) = \begin{cases} 0, & T_{fi} \leq T_{fi(1)}, \\ \frac{j}{n}, & T_{fi(j)} < T_{fi} \leq T_{fi(j+1)}, \\ 1, & T_{fi} > T_{fi(n)}; \end{cases}$$

$$F(T_{di}) = \begin{cases} 0, & T_{di} \leq T_{di(1)}, \\ \frac{j}{n}, & T_{di(j)} < T_{di} \leq T_{di(j+1)}, \\ 1, & T_{di} > T_{di(n)}; \end{cases}$$

$$F(S_i) = \begin{cases} 0, & S_i \leq S_{i(1)}, \\ \frac{j}{n}, & S_{i(j)} < S_i \leq S_{i(j+1)}, \\ 1, & S_i > S_{i(n)}; \end{cases}$$

$$F(Q_i) = \begin{cases} 0, & Q_i \leq Q_{i(1)}, \\ \frac{j}{n}, & Q_{i(j)} < Q_i \leq Q_{i(j+1)}, \\ 1, & Q_i > Q_{i(n)}; \end{cases}$$

где $T_{pi(j)}$, $T_{fi(j)}$, $T_{di(j)}$, $S_{i(j)}$, $Q_{i(j)}$ – j -е ранжированные значения параметров.

Далее запишем коэффициенты корреляции между временами, объемом и качеством разрабатываемых документов:

$$R(T_{pi}, S_i) = \frac{1}{n} \sum_{j=1}^n (T_{pij} - M[T_{pi}]) (S_{ij} - M[S_i]);$$

$$R(T_{pi}, Q_i) = \frac{1}{n} \sum_{j=1}^n (T_{pij} - M[T_{pi}]) (Q_{ij} - M[Q_i]);$$

$$R(T_{fi}, S_i) = \frac{1}{n} \sum_{j=1}^n (T_{fij} - M[T_{fi}]) (S_{ij} - M[S_i]);$$

$$R(T_{fi}, Q_i) = \frac{1}{n} \sum_{j=1}^n (T_{fij} - M[T_{fi}]) (Q_{ij} - M[Q_i]);$$

$$R(T_{di}, S_i) = \frac{1}{n} \sum_{j=1}^n (T_{dij} - M[T_{di}]) (S_{ij} - M[S_i]);$$

$$R(T_{di}, Q_i) = \frac{1}{n} \sum_{j=1}^n (T_{dij} - M[T_{di}]) (Q_{ij} - M[Q_i]);$$

$$R(S_i, Q_i) = \frac{1}{n} \sum_{j=1}^n (S_{ij} - M[S_i]) (Q_{ij} - M[Q_i]).$$

Регрессионные зависимости определяются методом наименьших квадратов:

$$T_{pi} = f(S_i) \approx A_0 + A_1 S_i + A_2 S_i^2 + \dots;$$

$$\sum_{j=1}^n (T_{pij} - A_0 + A_1 S_{ij} + A_2 S_{ij}^2 + \dots)^2 \rightarrow \min;$$

$$T_{fi} = f(S_i) \approx A_0 + A_1 S_i + A_2 S_i^2 + \dots;$$

$$\sum_{j=1}^n (T_{fij} - A_0 + A_1 S_{ij} + A_2 S_{ij}^2 + \dots)^2 \rightarrow \min;$$

$$T_{di} = f(S_i) \approx A_0 + A_1 S_i + A_2 S_i^2 + \dots;$$

$$\sum_{j=1}^n (T_{dij} - A_0 + A_1 S_{ij} + A_2 S_{ij}^2 + \dots)^2 \rightarrow \min;$$

$$T_{pi} = f(Q_i) \approx A_0 + A_1 Q_i + A_2 Q_i^2 + \dots;$$

$$\sum_{j=1}^n (T_{pij} - A_0 + A_1 Q_{ij} + A_2 Q_{ij}^2 + \dots)^2 \rightarrow \min;$$

$$T_{fi} = f(Q_i) \approx A_0 + A_1 Q_i + A_2 Q_i^2 + \dots;$$

$$\sum_{j=1}^n (T_{fij} - A_0 + A_1 Q_{ij} + A_2 Q_{ij}^2 + \dots)^2 \rightarrow \min;$$

$$T_{di} = f(Q_i) \approx A_0 + A_1 Q_i + A_2 Q_i^2 + \dots;$$

$$\sum_{j=1}^n (T_{dij} - A_0 + A_1 Q_{ij} + A_2 Q_{ij}^2 + \dots)^2 \rightarrow \min;$$

$$S_i = f(Q_i) \approx A_0 + A_1 Q_i + A_2 Q_i^2 + \dots;$$

$$\sum_{j=1}^n (S_{ij} - A_0 + A_1 Q_{ij} + A_2 Q_{ij}^2 + \dots)^2 \rightarrow \min.$$

Полученные результаты представляются пользователю для оптимальной организации работ по подготовке документов различного типа. В перспективе эту часть работы также можно автоматизировать на основе ма-

тематических моделей распределения ресурсов с оптимизацией выбранных показателей.

4 ПРИМЕР РАСЧЕТА

С использованием имитационного моделирования работы ДЛ органов военного управления с применением СЭД получены случайные значения времени подготовки документов и соответствующие им случайные значения объемов (табл. 1). На основе данных выборок рассчитаны коэффициенты регрессионных зависимостей. Графики зависимостей времени от объема приведены на рисунке 5.

Получены экспериментальные данные для построения аналитических зависимостей времени выполнения документов от их объема, которые необходимо использовать для оптимизации работ по формированию документов в СЭД.

ЗАКЛЮЧЕНИЕ

Важной задачей ДЛ органов управления является постоянное совершенствование своей деятельности путем организационных решений, распределения ресурсов, изменения алгоритма работы. Для этого необходимо контролировать процессы управления, в частности процессы разработки всех видов документов, выявляя проблемные места и пути их преодоления. Для частичной автоматизации такой работы предлагается расширить функциональные возможности ЭДО за счет включения подсистемы интеллектуального анализа

Рис. 5. Зависимости времени разработки документов $T(S)$ от их объема S

Таблица 1

Выборки данных при разработке документов разного типа

Документ 1			Документ 2			Документ 3		
№	S	T	№	S	T	№	S	T
1	17	31	1	34	68	1	41	104
2	15	30	2	30	66	2	35	91
3	10	27	3	21	59	3	21	64
4	17	31	4	35	69	4	43	109
5	13	29	5	27	64	5	31	83
6	11	28	6	22	60	6	23	67
7	19	32	7	39	70	7	48	120
8	17	31	8	35	69	8	42	106
9	18	31	9	36	69	9	44	110
10	15	30	10	31	67	10	37	95

Продолжение табл. 1

Документ 1			Документ 2			Документ 3		
№	S	T	№	S	T	№	S	T
11	17	31	11	35	69	11	42	106
12	14	29	12	29	65	12	34	89
13	16	30	13	33	68	13	40	102
14	11	28	14	23	61	14	24	69
15	13	29	15	27	64	15	30	81
16	13	29	16	26	63	16	30	81
17	16	30	17	32	67	17	39	100
18	10	27	18	21	59	18	22	66
19	14	29	19	29	65	19	34	89
20	19	32	20	39	70	20	49	123
$A_0 = 20; A_1 = 0,8; A_2 = -0,01$			$A_0 = 30; A_1 = 1,8; A_2 = -0,02$			$A_0 = 30; A_1 = 1,4; A_2 = 0,01$		

данных. Первый этап автоматизации заключается в установлении зависимостей между параметрами процессов разработки документов, на основе которых ДЛ должны принимать соответствующие решения. На последующих этапах необходимо формализовать принятие решений с использованием математических моделей.

СПИСОК ЛИТЕРАТУРЫ

1. Система автоматизации документооборота. URL: https://ru.wikipedia.org/wiki/Система_автоматизации_документооборота (дата обращения 06.01.2020).
2. Бобылева М.П. Управленческий документооборот: от бумажного к электронному. Вопросы теории и практики. М. : Термика, 2019. 470 с.
3. Системы электронного документооборота / Н.Ф. Алтухова, А.Л. Дзюбенко, В.В. Лосева, Ю.Б. Чечиков. М. : Кнорус, 2019. 202 с.
4. СЭД (рынок России). URL: [http://www.tadviser.ru/index.php/Статья:СЭД_\(рынок_России\)#](http://www.tadviser.ru/index.php/Статья:СЭД_(рынок_России)#) (дата обращения 08.01.2020).
5. «Умный» документооборот: как внедрять машинное обучение в СЭД и так ли это необходимо. URL: <http://www.tadviser.ru/a/417209> (дата обращения 06.01.2020).
6. Кучияш Е. TerraLink – о ключевых технологиях, которые будут менять ЕСМ-решения в ближайшем будущем. URL: http://www.tadviser.ru/index.php/Статья:Евгений_Кучияш_TerraLink:_Мы_наблюдаем_все_большее_вовлечение_наших_клиентов_в_гонку_цифровизации (дата обращения 11.01.2020).
7. Directum Ario. URL: http://www.tadviser.ru/index.php/Продукт:Directum_Ario# (дата обращения 11.01.2020).
8. Васильев Н.В., Забродин О.В., Куликов Д.В. Метод Process Mining для анализа регламентов документооборота // Техника средств связи. 2019. № 7 (146). С. 60–70.
9. Сбоева Л.И., Шамаева О.В. Базовые информационные технологии информационного тракта // Автоматизация процессов управления. 2006. № 2 (8). С. 65–71.

10. Круглов Б.В. Оптимальное распределение ресурсов при проектировании функциональных подсистем // Автоматизация процессов управления. 2011. № 2 (24). С. 3–7.

11. Круглов Б.В. Применение автоматизированных средств планирования в проектировании функциональных систем АСУ // Автоматизация процессов управления. 2012. № 2 (28). С. 9–16.

12. Иванов А.К. Математическое моделирование в технологии проектирования АСУ. Ульяновск, 2007. 270 с.

13. Иванов А.К., Типикин В.Г. Оптимизация разработки и реализации функциональных систем. Ульяновск, 2007. 207 с.

14. Методы и модели анализа данных: OLAP и DataMining : учеб. пособие / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. СПб, 2004. 336 с.

15. Справочник по вероятностным расчетам. М. : Воениздат, 1970. 536 с.

REFERENCES

1. *Sistema avtomatizatsii dokumentooborota* [Document Automation]. Available at: https://ru.wikipedia.org/wiki/Sistema_avtomatizatsii_dokumentooborota (accessed 06.01.2020).
2. Bobyleva M.P. *Upravlencheskii dokumentooborot: ot bumazhnogo k elektronnomu. Voprosy teorii i praktiki* [Document Management: Hard-to-Soft-Copy Document Workflow. Theoretical and Practical Issues]. Moscow, Termika Publ., 2019. 470 p.
3. Altukhova N.F., Dziubenko A.L., Loseva V.V., Chechikov Iu.B. *Sistemy elektronnoho dokumentooborota* [Electronic Document Management Systems]. Moscow, Knorus Publ., 2019. 202 p.
4. *SED (rynok Rossii)* [EDMS (Russian Market)]. Available at: [http://www.tadviser.ru/index.php/Stat'ya:SED_\(rynok_Rossii\)#](http://www.tadviser.ru/index.php/Stat'ya:SED_(rynok_Rossii)#) (accessed 08.01.2020).
5. *'Umnyi' dokumentooborot: kak vnedriat mashinnoe obuchenie v SED i tak li eto neobkhodimo* ['Smart' Document

Management: How to Implement the Machine Learning in EDMS, and is It Necessary?]. Available at: <http://www.tadviser.ru/a/417209> (accessed 06.01.2020).

6. Kuchiyash E. *TerraLink – o kliuchevykh tekhnologiyakh, kotorye budut meniat ECM-resheniya v blizhaishem budushchem* [TerraLink – Key Technologies, which will change the ECM-Solutions in the Nearest Future]. Available at: http://www.tadviser.ru/index.php/Stat'ya:Evgeniy_Kuchiyash_TerraLink:_My_nablyudaem_vse_bol'shee_vovlechenie_nashikh_klientov_v_gonku_tsifrovizatsii (accessed 11.01.2020).

7. *Directum Ario*. Available at: http://www.tadviser.ru/index.php/Produkt:Directum_Ario# (accessed 11.01.2020).

8. Vasilev N.V., Zabrodin O.V., Kulikov D.V. Metod 'Process Mining' dlia analiza reglamentov dokumentooborota [The 'Process Mining' Method for Analyzing the Document Management Regulations]. *Tekhnika sredstv svyazi* [Means of Communication Equipment], 2019, no. 7 (146). pp. 60–70.

9. Sboeva L.I., Shamaeva O.V. Bazovye informatsionnye tekhnologii informatsionnogo trakta [Basic Information Highway Technologies]. *Avtomatizatsiya protsessov upravleniya* [Automation of Control Processes], 2006, no. 2 (8), pp. 65–71.

10. Kruglov B.V. Optimalnoe raspredelenie resursov pri proektirovanii funktsionalnykh podsistem [Optimal Allocation of Resources for Design of Functional

Subsystems]. *Avtomatizatsiya protsessov upravleniya* [Automation of Control Processes], 2011, no. 2 (24), pp. 3–7.

11. Kruglov B.V. Primenenie avtomatizirovannykh sredstv planirovaniya v proektirovanii funktsionalnykh sistem ASU [Use of Computer-aided Planning Facilities for Design of Operational Systems of Computer-aided Control Systems]. *Avtomatizatsiya protsessov upravleniya* [Automation of Control Processes], 2012, no. 2 (28), pp. 9–16.

12. Ivanov A.K. *Matematicheskoe modelirovanie v tekhnologii proektirovaniya ASU* [Mathematical Modeling in the Procedure of C2 System Design]. Ulyanovsk, 2007. 270 p.

13. Ivanov A.K., Tipikin V.G. *Optimizatsiya razrabotki i realizatsii funktsionalnykh sistem* [The Optimization of Functional System Design and Implementation]. Ulyanovsk, 2007. 207 p.

14. Barsegian A.A., Kupriyanov M.S., Stepanenko V.V., Kholod I.I. *Metody i modeli analiza dannykh: OLAP i DataMining. Ucheb. posobie* [Data Mining Methods and Models: OLAP and Data Mining. Textbook]. St. Petersburg, 2004. 336 p.

15. *Spravochnik po veroiatnostnym raschetam* [Probability Calculation Guide]. Moscow, Voenizdat Pub., 1970. 536 p.