

ИНФОРМАТИКА,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И УПРАВЛЕНИЕ

INFORMATICS,
COMPPUTER ENGINEERING
AND CONTROL

УДК 004.048

DOI: 10.17212/2782-2001-2021-3-115-128

Разработка подсистемы интеллектуального анализа данных для системы электронного документооборота Citeck^{*}

Е.Е. ИСТРАТОВА^а, Д.Н. ДОСТОВАЛОВ^б

630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет

^а istratova@corp.nstu.ru ^б d.dostovalov@corp.nstu.ru

Актуальной задачей при внедрении систем электронного документооборота (СЭД) является расширение их функциональных возможностей за счет персонализации и учета индивидуальных особенностей организации. В статье рассматривается вопрос расширения функциональных возможностей системы электронного документооборота за счет проектирования подсистемы интеллектуального анализа данных. В рамках проведенного исследования были изучены принципы формализации процессов обработки входящей корреспонденции и организационно-распорядительных документов, способы сбора и анализа данных о работе пользователей с различными видами документов за счет применения искусственных нейронных сетей и комплексная оценка повышения эффективности работы СЭД образовательной организации. Для этого определены количественные характеристики, непосредственно влияющие на процесс контроля исполнения поручений, – временные затраты на создание и объем исполнения документа. На основе данных из системы электронного документооборота разработана математическая модель процесса создания документов, рассчитаны коэффициенты регрессии и получены аналитические зависимости качества разработанных документов от времени их исполнения и объема. Научная новизна исследования заключается в разработке алгоритма и программного обеспечения для автоматизации сбора и анализа данных за счет применения нейронных сетей в СЭД. К основным научным результатам относятся формализованные критерии документов и этапов их разработки, алгоритм работы подсистемы интеллектуального анализа данных и разработанное программное обеспечение для СЭД лица. Полученные результаты позволили выявить типы документации и этапы их разработки, отражающие наибольшие требования к ресурсам, необходимым для их выполнения, что в дальнейшем может быть использовано для поиска способов оптимальной организации работ по подготовке документов различного вида.

Ключевые слова: интеллектуальная система, электронный документооборот, Citeck, анализ данных, образовательная организация, машинное обучение, система электронного документооборота, жизненный цикл корреспонденции, оценка эффективности документооборота, количественные характеристики, качественные показатели

^{*} Статья получена 26 декабря 2020 г.

ВВЕДЕНИЕ

Необходимость в автоматизации процесса управления документооборотом практически в любой организации не вызывает сомнений с точки зрения эффективности управления ею. Предпосылками для этого являются следующие потребности предприятия:

- упорядочивание документооборота;
- повышение исполнительской дисциплины;
- формирование базы знаний организации;
- предоставление регламентированного доступа к необходимой информации.

В связи с интенсивным ростом информационных потоков в подразделениях образовательных организаций возникает потребность в применении комплексных средств автоматизации, которые смогли бы повысить оперативность, гибкость и мобильность принятия управленческих решений [1–4]. Так, переход от бумажного к электронному документообороту представляет собой одну из актуальных и первоочередных задач автоматизации образовательного учреждения [5–8]. Кроме того, в условиях усложнения бизнес-процессов образовательных учреждений для повышения их конкурентоспособности требуется усилить взаимодействие структурных подразделений с целью сокращения времени принятия управленческих решений [9–11]. Поэтому в качестве критериев для оптимизации деятельности подразделений выберем продолжительность и объем исполнения организационно-распорядительных документов.

При этом основная концепция внедряемых в настоящее время систем электронного документооборота исходит из следующих требований, предъявляемых к программным продуктам [12]:

- наличие веб-интерфейса для пользователей системы, причем как участников документооборота, так и разработчиков маршрутов документов, что объясняется необходимостью работы в системе большого числа пользователей, находящихся как в локальной сети образовательного учреждения, так и во внешней среде или даже в других городах региона [13–15];
- возможность автоматизированного управления доступом к документам системы в связи с большим числом часто меняющихся пользователей системы электронного документооборота;
- возможность масштабирования системы с точки зрения использования похожих маршрутов в различных структурах образовательной организации с возможностью разбивки по подразделениям;
- интеграция со всеми понятиями корпоративной информационной системы в связи с необходимостью реализации бизнес-процессов всех направлений деятельности учреждения [16];
- возможность реализации различных маршрутов документов, а также маршрутов работы [17].

Исходя из этого использование системы электронного документооборота в образовательной организации позволяет не только контролировать потоки информации, но и значительно сокращать сроки обработки документов, что в результате дает возможность оптимизировать процессы принятия управленческих решений. Согласно литературным источникам [1, 3, 18–21], опыт внедрения систем электронного документооборота позволил решить задачи, связанные как со скоростью обработки информации, так и с защитой обрабатываемых и передаваемых данных [22, 23].

Таким образом, формирование документооборота в современном образовательном учреждении в качестве базового структурного элемента позволяет повысить эффективность работы образовательной организации, провести корректировку бизнес-процессов, обеспечить возможность использования данных в соответствии с нормативными требованиями и ускорить обработку документов. В результате внедрение в эксплуатацию системы электронного документооборота способно оказать положительное влияние на повышение конкурентоспособности образовательного учреждения в целом.

Несмотря на доказанные преимущества от внедрения в эксплуатацию систем электронного оборота, в некоторых случаях требуются дополнительные информационные технологии и инструменты для анализа качества обработки данных. В первую очередь это связано с интенсивным ростом объемов информации в сфере образования и необходимостью принятия оперативных управленческих решений. Помимо этого, подобный анализ значительно упрощает контроль выполнения поручений и может быть применен как метод повышения исполнительской дисциплины. Для реализации данного направления целесообразно использовать методы интеллектуального анализа данных [24–27].

Целью исследования являлось расширение функциональных возможностей системы электронного документооборота образовательной организации за счет проектирования подсистемы интеллектуального анализа данных.

Исследование проводилось на базе образовательного учреждения – лицея № 22 «Надежда Сибири» города Новосибирска. В качестве исходных данных были рассмотрены входящие, исходящие и внутренние документы организации, а также взаимосвязи между ними. Основные задачи разработки подсистемы интеллектуального анализа данных:

- исследование этапов работы с различными типами документов;
- структурирование документов по функциональному признаку;
- определение критериев, влияющих на скорость и качество обработки документов;
- выявление зависимостей между данными критериями и эффективностью организации документооборота предприятия в целом;
- анализ результатов и их применение при разработке подсистемы интеллектуального анализа данных.

Поскольку проектируемая подсистема интеллектуального анализа данных является одним из элементов системы электронного документооборота организации, в ее функции должны входить такие операции:

- сбор данных об особенностях и временных характеристиках разработки документов;
- определение вероятностных характеристик на основе собранных данных;
- выявление статистических зависимостей (корреляция, регрессия) между различными критериями разработки организационно-распорядительных документов;
- анализ полученных результатов для дальнейшего формирования решений по оптимизации процесса организации работ с документами.

1. ХАРАКТЕРИСТИКА СИСТЕМЫ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА ЛИЦЕЯ

Совокупность документов рассматриваемой образовательной организации можно условно разделить на две группы: организационно-распорядительная документация и корреспонденция. На основе собранных данных об

организационной структуре лица № 22 «Надежда Сибири» и о существующих бизнес-процессах была спроектирована блок-схема, описывающая жизненный цикл входящей корреспонденции (рис. 1).

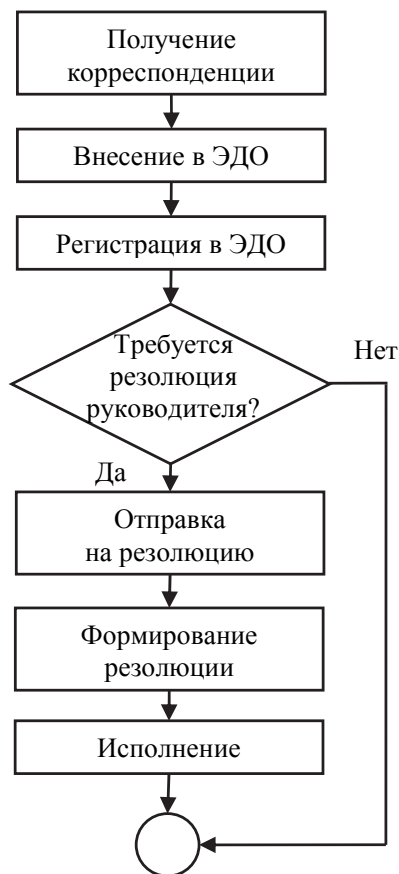


Рис. 1. Жизненный цикл входящей корреспонденции

Fig. 1. Life cycle of incoming documents

Жизненный цикл входящей корреспонденции начинается в лице с ее поступления и внесения в систему электронного документооборота. Следующим принципиальным этапом является регистрация документа, осуществляемая путем внесения отдельных его реквизитов в соответствующие поля. После регистрации необходимо решить, требуется ли на конкретном документе резолюция руководителя, т. е. директора лица. Если резолюция не требуется, документ отправляется далее по маршруту на исполнение, а потом в архив, если не предусмотрено других вариантов. Если резолюция требуется, то сначала документ направляется на резолюцию, после формирования которой он идет на исполнение.

В ходе исследования были определены основные этапы утверждения организационно-распорядительных документов в организации. К подобным документам были отнесены приказы, инструкции, правила и т. д. Все документы данного вида готовятся, оформляются и исполняются в пределах самой

организации. Основной порядок их прохождения и утверждения включает следующие этапы: согласование, подписание, ознакомление, начало действия документа. Кроме перечисленных основных этапов, маршрут следования документа также может включать еще несколько стадий (например, доработка, сканирование). На этапе подписания документ может отправиться на доработку и опять вернуться на подпись, также документ может проходить несколько раз эту процедуру, прежде чем его отправят на ознакомление персоналу. Подробная диаграмма всех этапов утверждения организационно-распорядительной документации приведена на рис. 2.

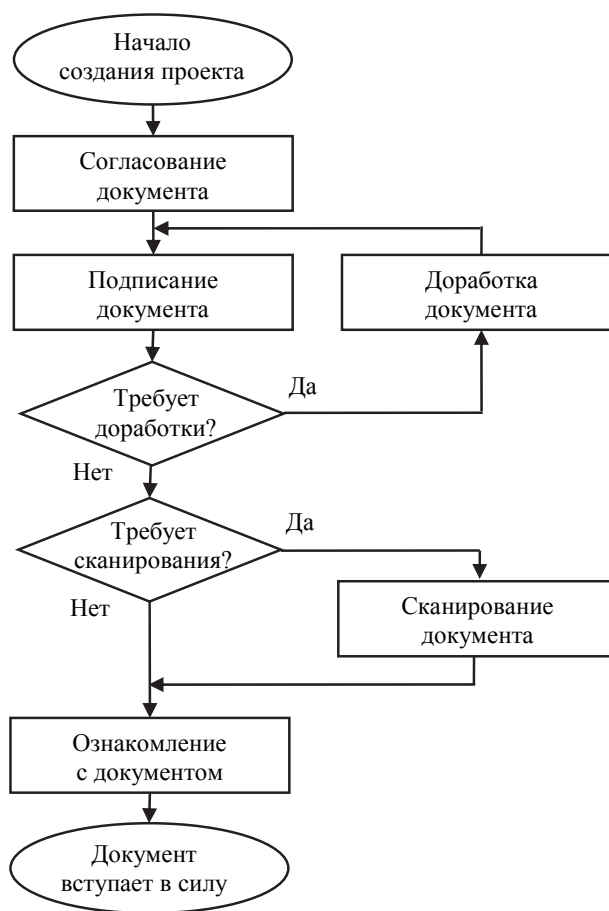


Рис. 2. Диаграмма этапов утверждения документа

Fig. 2. Diagram of document approval stages

Эксплуатация системы электронного документооборота образовательного учреждения включает в себя автоматизацию перемещения основных информационных потоков документов между исполнителями и утверждающими должностными лицами. При этом обмен всеми данными, к которым относятся как генерируемые системой сообщения, так и файлы пользователей, осуществляется посредством системы без непосредственного общения между участниками процесса.

2. ПОЛУЧЕНИЕ ДАННЫХ ИЗ СИСТЕМЫ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА

В исследовании предлагается расширение функциональных возможностей системы электронного документооборота образовательной организации за счет проектирования подсистемы интеллектуального анализа данных. Ее основными задачами являются: сбор информации о процессах разработки и согласования документов; определение вероятностных характеристик и зависимостей между различными показателями, влияющими на процесс создания документов; вывод результатов анализа для принятия пользователем решения по организации документооборота. В результате возможно сформировать совокупность правил и выстроить приоритеты при передаче и обработке документов.

Для реализации указанных задач подсистема интеллектуального анализа данных должна содержать следующие три функциональных модуля, отвечающие за планирование, разработку и контроль выполнения документов соответственно. Состав модулей подсистемы интеллектуального анализа данных приведен на рис. 3.

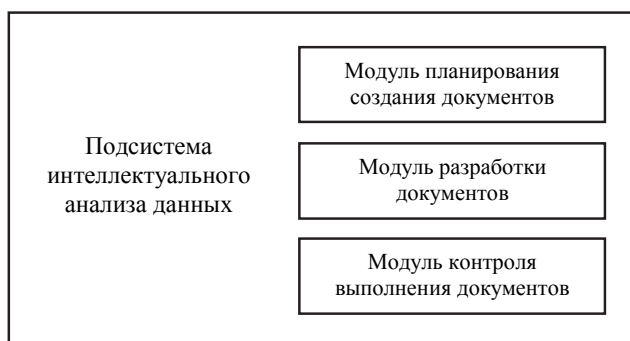


Рис. 3. Структура подсистемы интеллектуального анализа данных

Fig. 3. Structure of the data mining subsystem

На этапе планирования создания документа пользователь, инициирующий этот процесс, с помощью инструментария существующей системы электронного документооборота разрабатывает план формирования документа, назначает ответственных исполнителей, а также в зависимости от ситуации добавляет график и определяет время, необходимое для разработки документа. В качестве основы для нового документа может быть использован шаблон ранее разработанного документа, к которому добавляются пользователи-исполнители. Проектируемый шаблон документа отображается в виде диаграммы Ганта. В рамках этого модуля осуществляется составление плана-графика изготовления документа, производится сбор данных об исполнителях, маршруте следования и составе полей документа.

Процесс разработки документа начинается с момента его запуска инициатором процесса, имеющим соответствующие права. При этом в процессе запуска каждый документ проходит проверку на корректность в соответствии с определенными установленными критериями. Если документ содержит ошиб-

ки, то у пользователя отображается сообщение об ошибке и список всех некорректных элементов. После успешного выполнения документа ответственные исполнители получают соответствующее оповещение и дальше могут приступать к оформлению, отправке на согласование и утверждение документов за счет средств электронного документооборота организации. В рамках этого модуля осуществляется сбор данных о времени изготовления документа и выполнении запланированных разделов и показателей качества.

На этапе контроля выполнения документов пользователи, задействованные в процессах разработки или согласования конкретного документа, могут отслеживать ход его выполнения, т. е. имеют доступ к графику и статистическим данным по пользователям-исполнителям. Такой модуль позволяет собрать и проанализировать информацию о выполнении документа, его соответствии изначально запланированным характеристикам и срокам.

Таким образом, использование подсистемы интеллектуального анализа данных дает возможность выявить проблемные виды и этапы создания документов, необходимые для поиска как методов оптимизации документооборота в целом, так и путей повышения качества отдельных документов.

3. ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ЭЛЕКТРОННЫХ ДОКУМЕНТОВ

Для развития существующей системы электронного документооборота предлагается применить методы интеллектуального анализа данных, позволяющие реализовать функциональную подсистему на основе выявленных математических зависимостей между критериями разработки и учета различных видов документов [8, 9, 14].

В данном исследовании под интеллектуальным анализом данных подразумевается как поиск и объяснение закономерностей в получаемых из системы электронного документооборота образовательной организации данных, так и их вывод в разрезе эффективности работы структурных подразделений лицея над отдельными видами документов.

Таким образом, проектирование подсистемы интеллектуального анализа данных на основе существующей системы электронного документооборота образовательной организации позволило выявить виды документов и этапы их разработки, наиболее требовательные к ресурсам, необходимым для их выполнения, что, в свою очередь, дает возможность поиска способов для ускорения разработки документов и повышения их качества.

Сведения, содержащие информацию о разработке, согласовании и утверждении различных видов документов, хранятся в базе данных системы электронного документооборота образовательной организации. Однако для проведения интеллектуального анализа данных необходимо дополнить уже существующий набор двумя количественными характеристиками, непосредственно влияющими на процесс контроля исполнения поручений. Такими характеристиками являются показатели оперативности и объема исполнения документов.

Первый количественный критерий напрямую связан с затратами времени на подготовку конкретного документа строго определенного вида, относящегося к организационно-распорядительной документации или к корреспонденции. Этот критерий показывает отношение времени, затраченного на

разработку документа, к запланированному времени. В математическом виде зависимость процесса разработки документа от времени можно представить следующим образом:

$$T_d = \frac{T_{\Phi}}{T_{\Pi}}, \quad (1)$$

$$T_d = \{T_{d1}, T_{d2}, \dots, T_{dn}\}, \quad (2)$$

где T_d – критерий, характеризующий временные затраты на создание документа; T_{Φ} – время, фактически затраченное на изготовление документа; T_{Π} – время, запланированное на создание документа; n – количество данных в выборке.

Рассматриваемые документы могут быть разделены на виды и этапы выполнения, что дает возможность анализа по каждому виду отдельно. С учетом этого формула (2) примет вид

$$T_{dij} = \{T_{dij1}, T_{dij2}, \dots, T_{dijn}\}, \quad (3)$$

$$i = [1, m], \quad j = [1, k],$$

где i – номер вида документа; m – количество видов документов; j – номер этапа разработки документа; k – количество этапов.

Второй количественный критерий для проведения интеллектуального анализа данных связан с объемом исполнения документов. Так как на величину этого критерия влияет совокупность различных факторов, то для упрощения математической модели учитывается только объем разработанных и при необходимости утвержденных документов. В математическом виде зависимость процесса разработки документа от объема его выполнения можно представить следующим образом:

$$V_d = \frac{V_{\Phi}}{100 \%}, \quad (4)$$

$$V_{dij} = \{V_{dij1}, V_{dij2}, \dots, V_{dijn}\}, \quad (5)$$

где V_d – критерий, характеризующий объем исполнения документа; V_{Φ} – фактическая доля от объема исполнения документа, %; n – количество данных в выборке; i – номер вида документа; m – количество видов документов; j – номер этапа разработки документа; k – количество этапов.

В качестве результирующей функции оценки выполнения заданного регламента целесообразно рассмотреть качество разработанного документа Q , определяемое по формуле

$$Q = T_{dij} V_{dij}, \quad (6)$$

С точки зрения качества изготовления документа его объем выполнения и скорость разработки имеют одинаковые приоритеты, т. е. при высокой скорости и недостаточном объеме исполнения итоговая интегральная оценка качества будет невысокой. При необходимости можно получить отчет по каж-

дому из критериев отдельно. Для интерпретации уровня качества документов были установлены интервалы (табл. 1).

Таблица 1

Table 1

Интерпретация значений Q

Interpretation of Q values

Интервал значений Q	Оценка уровня качества
Менее 0,3	Неудовлетворительно
0,3...0,5	Удовлетворительно
0,5...0,7	Хорошо
0,7...1,0	Отлично

Проведенный в рамках исследования интеллектуальный анализ данных включал в себя следующие этапы:

- определение цели и задач анализа;
- сбор и предварительная обработка данных, необходимых для проведения анализа;
- построение моделей за счет использования методологии Data Mining;
- проверка полученных математических моделей;
- интерпретация результатов.

На основе данных, полученных из системы электронного документооборота лица, была разработана математическая модель, описывающая процесс создания документов в образовательной организации, были рассчитаны коэффициенты регрессии и получены экспериментальные значения для построения аналитических зависимостей качества разработанных документов от времени их исполнения и объема.

В качестве исходных данных для проведения анализа качества и скорости разработки документов были рассмотрены организационно-распорядительные документы и входящая корреспонденция образовательной организации. Предварительно документы были разделены на восемь типов, для каждого из которых были определены качественные показатели, рассчитаны коэффициенты регрессии и получены экспериментальные значения для построения аналитических зависимостей качества разработанных документов от времени их исполнения и объема (табл. 2).

Таблица 2

Table 2

Результаты анализа

Results of the analysis

Документы	T	V	Q
Тип 1	0,73	0,79	0,58
Тип 2	0,84	0,93	0,78
Тип 3	0,96	0,84	0,81
Тип 4	0,84	0,91	0,76
Тип 5	0,69	0,83	0,57
Тип 6	0,91	0,84	0,76
Тип 7	0,94	0,89	0,84
Тип 8	0,97	0,95	0,92

На основе такого набора данных построено восемь математических моделей, качественные показатели шести из которых располагались в интервале от 0,7 до 1,0.

ЗАКЛЮЧЕНИЕ

Полученные результаты позволили выявить, что наиболее требовательными к ресурсам являются такие виды документов, как внутренние распоряжения и входящая корреспонденция, причем если первый вид документов значительно отстает на этапе разработки, то второй – на этапе планирования, т. е. при регистрации, что существенно влияет на итоговый качественный показатель. Выявленные закономерности могут быть использованы для поиска способов оптимальной организации работы по подготовке документов различного вида с учетом различных этапов работы.

Таким образом, в результате проведения исследования была разработана и протестирована на реальных значениях подсистема интеллектуального анализа данных, позволяющая расширить функциональные возможности существующей в образовательной организации системы электронного документооборота для дальнейшего формирования решений по оптимизации процесса организации работ с документами.

СПИСОК ЛИТЕРАТУРЫ

1. Системы электронного документооборота / Н.Ф. Алтухова, А.Л. Дзюбенко, В.В. Лосева, Ю.Б. Чечиков. – М.: Кнорус, 2019. – 202 с.
2. Истратова Е.Е., Син Д.Д. Выбор Data Mining-инструмента для обработки научных данных // Вызовы цифровой экономики: развитие комфортной городской среды: сборник статей III Всероссийской научно-практической конференции с международным участием. – Брянск, 2020. – С. 348–352.
3. Черников А.А., Истратова Е.Е. Применение систем электронного документооборота в сфере малого бизнеса города Новосибирска // Актуальные проблемы мировой экономики и менеджмента: материалы международной интернет-конференции. – Гомель, 2019. – С. 256–257.
4. Симонова С.И. Интеллектуальный анализ данных для задач CRM // International Journal of Open Information Technologies. – 2015. – Vol. 3, N 2. – URL: <https://cyberleninka.ru/article/n/intellektualnyy-analiz-dannyh-dlya-zadach-crm> (accessed: 01.09.2021).
5. Ефремова Л.И., Колекина А.О. Выбор системы электронного документооборота для предприятия // Вестник Волжского университета им. В.Н. Татищева. – 2019. – Т. 2, № 1. – URL: <https://cyberleninka.ru/article/n/vybor-sistemy-elektronnogo-dokumentooborota-dlya-predpriyatiya> (дата обращения: 01.09.2021).
6. Апришко Д.В., Таран В.Н. Системы электронного документооборота и их обзор // Дистанционные образовательные технологии: материалы IV Всероссийской научно-практической конференции (с международным участием). – Симферополь, 2019. – С. 244–249. – URL: <https://www.elibrary.ru/item.asp?id=40453380> (дата обращения: 01.09.2021).
7. Lange P. de, Nicolaescu P., Neumann A.T. Integrating web-based collaborative live editing and wireframing into a model-driven web engineering process // Data Science and Engineering. – 2020. – Vol. 5. – P. 240–260. – DOI: 10.1007/s41019-020-00131-3.
8. Медведева О.В., Парамонова М.Г. Цифровизация управления и системы электронного документооборота // Ученые записки Тамбовского отделения РoCMY. – 2019. – № 13. – URL: <https://cyberleninka.ru/article/n/tsifrovizatsiya-upravleniya-i-sistemy-elektronnogo-dokumentooborota> (дата обращения: 02.09.2021).
9. Patil T., Bhavsar A.K. Data science team roles and need of data science: a review of different cases // Data Science and Intelligent Applications: Proceedings of ICDSIA 2020. – Singapore:

Springer, 2020. – (Lecture Notes on Data Engineering and Communications Technologies; vol. 52). – DOI: 10.1007/978-981-15-4474-3_2.

10. *Bhagchandani A., Trivedi D.* a machine learning algorithm to predict financial investment // Data Science and Intelligent Applications: Proceedings of ICDSIA 2020. – Singapore: Springer, 2020. – (Lecture Notes on Data Engineering and Communications Technologies; vol. 52). – DOI: 10.1007/978-981-15-4474-3_30.

11. *Paramonova I.E.* Electronic document-management systems: a classification and new opportunities for a scientific technical library // Scientific and Technical Information Processing. – 2016. – Vol. 43. – P. 136–143. – DOI: 10.3103/S0147688216030047.

12. *Ginsburg M.* An agent framework for intranet document management // Autonomous Agents and Multi-Agent Systems. – 1999. – Vol. 2. – P. 271–286. – DOI: 10.1023/A:1010012406205.

13. *Hillah L.M., Maesano A.P., Rosa F. De.* Automation and intelligent scheduling of distributed system functional testing // International Journal on Software Tools for Technology Transfer. – 2017. – Vol. 19. – P. 281–308. – DOI: 10.1007/s10009-016-0440-3.

14. *Ahmad K., Sahu M., Shrivastava M.* An efficient image retrieval tool: query based image management system // International Journal of Information Technology. – 2020. – Vol. 12. – P. 103–111. – DOI: 10.1007/s41870-018-0198-9.

15. *Kleshchev A.S., Chernyakhovskaya M.Y., Shalfeeva E.A.* Features of the automation of intellectual activities // Automatic Documentation and Mathematical Linguistics. – 2015. – Vol. 49. – P. 10–20. – DOI: 10.3103/S0005105515010021.

16. *Kolekar S., Sanjeevi S., Bormane D.S.* The framework of an adaptive user interface for e-learning environment using artificial neural network // International Conference on E-Learning E-Business, IEEE 2010. – Las Vegas, Nevada, USA, 2010. – P. 65–69.

17. *Fernandez-Garca A.J.* A recommender system for component-based applications using machine learning techniques // Knowledge-Based Systems. – 2019. – Vol. 164. – P. 68–84. – DOI: 10.1016/j.knosys.2018.10.019.

18. *Okuda H., Ogata S., Matsuura S.* Experimental development based on mapping rule between requirements analysis model and web framework specific design model // SpringerPlus. – 2013. – Vol. 2. – P. 123. – DOI: 10.1186/2193-1801-2-123.

19. *Бобылева М.П.* Управленческий документооборот: от бумажного к электронному: вопросы теории и практики. – М.: Термика, 2019. – 470 с.

20. *Obukhov A., Krasnyanskiy M., Nikolyyukin M.* Algorithm of adaptation of electronic document management system based on machine learning technology // Progress in Artificial Intelligence. – 2020. – Vol. 9. – P. 287–303. – DOI: 10.1007/s13748-020-00214-2.

21. *Ghaibi N., Dassi O., Ayed L.* User interface adaptation based on a business rules management system and machine learning // Communications of the IBIMA. – 2018. – Vol. 2018. – Art. 281881. – DOI: 10.5171/2018.281881.

22. Model-based frameworks for user adapted information exploration: an overview / M. Kotzyba, T. Gossen, S. Stober, A. Nurnberger // Companion Technology. – Cham: Springer, 2017. – P. 37–56. – DOI: 10.1007/978-3-319-43665-4_3.

23. Multi-level preference regression for cold-start recommendations / F. Peng, X. Lu, C. Ma, Y. Qian, J. Lu, J. Yang // International Journal of Machine Learning and Cybernetics. – 2018. – Vol. 9, N 7. – P. 1117–1130.

24. *Кривенко Ю.С., Минасян А.Т., Разиньков А.О.* Исследование технологий интеллектуального анализа данных (Data Mining) // Актуальные проблемы управления в электронной экономике. Одиннадцатые Ходыревские чтения. – Курск, 2018. – С. 182–184.

25. *Elmabaredy A., Elkholy E., Tolba A.A.* Web-based adaptive presentation techniques to enhance learning outcomes in higher education // Research and Practice in Technology Enhanced Learning. – 2020. – Vol. 15. – DOI: 10.1186/s41039-020-00140-w.

26. *Шокин Ю.И., Юрченко А.В.* О моделях организации хранения и использования научных данных: основные принципы, процессы и механизмы // Информационно-управляющие системы. – 2019. – № 3. – С. 45–54.

27. Методы и модели анализа данных: OLAP и DataMining: учебное пособие / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – СПб.: БХВ-Петербург, 2004. – 336 с.

Истратова Евгения Евгеньевна, кандидат технических наук, доцент кафедры автоматизированных систем управления Новосибирского государственного технического университета. Основное направление научных исследований – сетевые информационные технологии, информационные системы, интеллектуальный анализ данных. Имеет более 50 печатных работ и учебных пособий. E-mail: istratova@corp.nstu.ru

Достовалов Дмитрий Николаевич, кандидат технических наук, доцент кафедры автоматизированных систем управления Новосибирского государственного технического университета. Основное направление научных исследований – математическое и программное обеспечение, компьютерное моделирование, сетевые информационные технологии. Имеет более 70 печатных работ и учебных пособий. E-mail: d.dostovalov@corp.nstu.ru

Istratova Evgeniya E., PhD (Eng.), associate professor at the Department of Automated Control Systems, Novosibirsk State Technical University. The main field of her research is network information technologies, information systems, and data mining. She has over 50 publications and tutorials. E-mail: istratova@corp.nstu.ru

Dostovalov Dmitry N., PhD (Eng.), associate professor at the Department of Automated Control Systems, Novosibirsk State Technical University. The main field of his research is mathematical support and software, computer modeling, and network information technologies. He has over 70 publications and tutorials. E-mail: d.dostovalov@corp.nstu.ru

DOI: 10.17212/2782-2001-2021-3-115-128

Development of a data mining subsystem for the Citeck electronic document management system*

E.E. ISTRATOVA^a, D.N. DOSTOVALOV^b

Novosibirsk State Technical University, 20 K. Marx Prospekt, Novosibirsk, 630073, Russian Federation

^a istratova@corp.nstu.ru ^b d.dostovalov@corp.nstu.ru

Abstract

An urgent task in the implementation of electronic document management systems (EDMS) is to expand their functionality through personalization and taking into account individual characteristics of the organization. The article deals with expanding the functionality of EDMS by designing a subsystem for data mining. As part of the study, the principles of formalizing the processes of processing incoming correspondence and organizational and administrative documents, methods of collecting and analyzing data on the work of users with various types of documents through the use of artificial neural networks and a comprehensive assessment of improving the efficiency of the EDMS of an educational institution were studied. Quantitative characteristics that directly affect the process of monitoring the execution of orders have been determined. This is the time spent for creation of the document and execution completeness of the document. A mathematical model of the process of creating documents based on data from the EDMS has been developed. Regression coefficients have been calculated. Analytical dependences of the quality of the developed documents on the time of their execution and volume have been obtained. The scientific novelty of the research lies in the development of an algorithm and software for automating the collection and analysis of data through the use of neural networks in the EDMS. The main scientific results include formalized criteria for documents and stages of their development, the algorithm of the mining subsystem, the developed software for the EDMS of the Lyceum. The results obtained made it possible to identify the types of documents and the stages of their development that are most demanding on the resources necessary for their implementation, which can later be used to find ways to optimally organize work on the preparation of documents of various types.

* Received 26 December 2020.

Keywords: intelligent system, electronic document management, Citeck, data analysis, educational organization, machine learning, electronic document management system

REFERENCES

1. Altukhova N.F., Dzyubenko A.L., Loseva V.V., Chechikov Yu.B. *Sistemy elektronnoho dokumentooborota* [Electronic document management systems]. Moscow, Knorus Publ., 2019. 202 p.
2. Istratova E.E., Sin D.D. [The data mining tools selection for scientific data processing]. *Vyzovy tsifrovoy ekonomiki: razvitie komfortnoi gorodskoi sredy: sbornik statei III Vserossiiskoi nauchno-prakticheskoi konferentsii s mezhdunarodnym uchastiem* [Proceedings of the III All-Russian Scientific and Practical Conference "Challenges of the Digital Economy: Development of a Comfortable Urban Environment"], Bryansk, 2020, pp. 348–352. (In Russian).
3. Chernikov A.A., Istratova E.E. [Using of electronic document management systems in the small business of the city of Novosibirsk]. *Aktual'nye problemy mirovoi ekonomiki i menedzhmenta: materialy mezhdunarodnoi internet-konferentsii* [Proceedings of the International Internet conference "Actual problems of the world economy and management"], Gomel, 2019, pp. 256–257. (In Russian).
4. Simonova S.I. Intellektual'nyi analiz dannykh dlya zadach CRM [Data mining for CRM tasks]. *International Journal of Open Information Technologies*, 2015, vol. 3, no. 2. (In Russian). Available at: <https://cyberleninka.ru/article/n/intellektualnyy-analiz-dannyh-dlya-zadach-crm> (accessed 01.09.2021).
5. Efremova L.I., Kolekina A.O. Vybory sistem elektronnoho dokumentooborota dlya predpriyatiya [Choice of electronic document turning system for enterprise]. *Vestnik Volzhskogo universiteta im. V.N. Tatishcheva = Vestnik of Volzhsky University after V.N. Tatishchev*, 2019, vol. 2, no. 1. Available at: <https://cyberleninka.ru/article/n/vybor-sistemy-elektronnoho-dokumentooborota-dlya-predpriyatiya> (accessed 01.09.2021).
6. Aprishko D.V., Taran V.N. [Electronic document management systems and their review]. *Distantionnye obrazovatel'nye tekhnologii: materialy IV Vserossiiskoi nauchno-prakticheskoi konferentsii (s mezhdunarodnym uchastiem)* [Distance educational technologies. Proceedings of the IV All-Russian Scientific and Practical Conference (with international participation)]. Simferopol, 2019, pp. 244–249. (In Russian). Available at: <https://www.elibrary.ru/item.asp?id=40453380> (accessed 01.09.2021).
7. Lange P. de, Nicolaescu P., Neumann A.T. Integrating web-based collaborative live editing and wireframing into a model-driven web engineering process. *Data Science and Engineering*, 2020, vol. 5, pp. 240–260. DOI: 10.1007/s41019-020-00131-3.
8. Medvedeva O.V., Paramonova M.G. Tsifrovizatsiya upravleniya i sistemy elektronnoho dokumentooborota [Digitalization of control and electronic system document flow]. *Uchenye zapiski Tambovskogo otdeleniya RoSMU = Scientific notes of the Tambov branch of the RUYS*, 2019, no. 13. Available at: <https://cyberleninka.ru/article/n/tsifrovizatsiya-upravleniya-i-sistemy-elektronnoho-dokumentooborota> (accessed 02.09.2021).
9. Patil T., Bhavsar A.K. Data science team roles and need of data science: a review of different cases. *Data Science and Intelligent Applications: Proceedings of ICDSIA 2020. Lecture Notes on Data Engineering and Communications Technologies*, vol. 52. Singapore, Springer, 2020. DOI: 10.1007/978-981-15-4474-3_2.
10. Bhagchandani A., Trivedi D. A machine learning algorithm to predict financial investment. *Data Science and Intelligent Applications: Proceedings of ICDSIA 2020. Lecture Notes on Data Engineering and Communications Technologies*, vol. 52. Singapore, Springer, 2020. DOI: 10.1007/978-981-15-4474-3_30.
11. Paramonova I.E. Electronic document-management systems: a classification and new opportunities for a scientific technical library. *Scientific and Technical Information Processing*, 2016, vol. 43, pp. 136–143. DOI: 10.3103/S0147688216030047.
12. Ginsburg M. An agent framework for intranet document management. *Autonomous Agents and Multi-Agent Systems*, 1999, vol. 2, pp. 271–286. DOI: 10.1023/A:1010012406205.
13. Hillah L.M., Maesano A.P., Rosa F. De. Automation and intelligent scheduling of distributed system functional testing. *International Journal on Software Tools for Technology Transfer*, 2017, vol. 19, pp. 281–308. DOI: 10.1007/s10009-016-0440-3.
14. Ahmad K., Sahu M., Shrivastava M. An efficient image retrieval tool: query based image management system. *International Journal of Information Technology*, 2020, vol. 12, pp. 103–111. DOI: 10.1007/s41870-018-0198-9.

15. Kleshchev A.S., Chernyakhovskaya M.Y., Shalfeeva E.A. Features of the automation of intellectual activities. *Automatic Documentation and Mathematical Linguistics*, 2015, vol. 49, pp. 10–20. DOI: 10.3103/S0005105515010021.
16. Kolekar S., Sanjeevi S., Bormane D.S. The framework of an adaptive user interface for e-learning environment using artificial neural network. *International Conference on E-Learning E-Business, IEEE 2010*, Las Vegas, Nevada, USA, 2010, pp. 65–69.
17. Fernandez-Garcia A.J. A recommender system for component-based applications using machine learning techniques. *Knowledge-Based Systems*, 2019, vol. 164, pp. 68–84. DOI: 10.1016/j.knosys.2018.10.019.
18. Okuda H., Ogata S., Matsuura S. Experimental development based on mapping rule between requirements analysis model and web framework specific design model. *SpringerPlus*, 2013, vol. 2, p. 123. DOI: 10.1186/2193-1801-2-123.
19. Bobyleva M.P. *Upravlencheskii dokumentooborot: ot bumazhnogo k elektronnomu: vo-prosy teorii i praktiki* [Management document flow: from paper to electronic]. Moscow, Termika Publ., 2019. 470 p.
20. Obukhov A., Krasnyanskiy M., Nikolyukin M. Algorithm of adaptation of electronic document management system based on machine learning technology. *Progress in Artificial Intelligence*, 2020, vol. 9, pp. 287–303. DOI: 10.1007/s13748-020-00214-2.
21. Ghaibi N., Dassi O., Ayed L. User interface adaptation based on a business rules management system and machine learning. *Communications of the IBIMA*, 2018, vol. 2018, art. 281881. DOI: 10.5171/2018.281881.
22. Kotzyba M., Gossen T., Stober S., Nurnberger A. Model-based frameworks for user adapted information exploration: an overview. *Companion Technology*. Cham, Springer, 2017, pp. 37–56. DOI: 10.1007/978-3-319-43665-4_3.
23. Peng F., Lu X., Ma C., Qian Y., Lu J., Yang J. Multi-level preference regression for cold-start recommendations. *International Journal of Machine Learning and Cybernetics*, 2018, vol. 9, no. 7, pp. 1117–1130.
24. Krivenko Yu.S., Minasyan A.T., Razin'kov A.O. [Research of technologies of Data Mining]. *Aktual'nye problemy upravleniya v elektronnoi ekonomike. Odinnadtsatye Khodyrevskie chteniya* [Actual problems of management in the electronic economy. Eleventh Khodyrev Readings], Kursk, 2018, pp. 182–184. (In Russian).
25. Elmabaredy A., Elkholy E., Tolba A.A. Web-based adaptive presentation techniques to enhance learning outcomes in higher education. *Research and Practice in Technology Enhanced Learning*, 2020, vol. 15. DOI: 10.1186/s41039-020-00140-w.
26. Shokin Yu.I., Yurchenko A.V. O modelyakh organizatsii khraneniya i ispol'zovaniya nauchnykh dannykh: osnovnye printsipy, protsessy i mekhanizmy [On the models for organizing the storage and use of scientific data: basic principles, processes and mechanisms]. *Informatsionno-upravlyayushchie sistemy = Information and Control Systems*, 2019, no. 3, pp. 45–54.
27. Barsegyan A.A., Kupriyanov M.S., Stepanenko V.V., Kholod I.I. *Metody i modeli analiza dannykh: OLAP i DataMining* [Methods and models of data analysis: OLAP and DataMining]. St. Petersburg, BHV-Peterburg Publ., 2004. 336 p.

Для цитирования:

Истратова Е.Е., Достовалов Д.Н. Разработка подсистемы интеллектуального анализа данных для системы электронного документооборота Citeck // Системы анализа и обработки данных. – 2021. – № 3 (83). – С. 115–128. – DOI: 10.17212/2782-2001-2021-3-115-128.

For citation:

Istratova E.E., Dostovalov D.N. Razrabotka podsistemy intellektual'nogo analiza dannykh dlya sistemy elektronnoy dokumentooborota Citeck [Development of a data mining subsystem for the Citeck electronic document management system]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2021, no. 3 (83), pp. 115–128. DOI: 10.17212/2782-2001-2021-3-115-128.