

Боцвин Сергей Александрович – Главное управление кадров Министерства обороны Российской Федерации (г. Москва); e-mail: xvatkoff@mail.ru; г. Москва, Россия, тел.: 89856834525; зам. начальника.

Хватков Вадим Александрович – Военный учебный центр при Национальном исследовательском университете «Московский институт электронной техники» (г. Москва); e-mail: xvatkoff@mail.ru; г. Москва, г. Зеленоград, Россия; тел.: 89264608261; начальник учебной части – заместитель начальника.

Botsvin Sergey Aleksandrovich – The Main personnel Directorate of the Ministry of defense of the Russian Federation (Moscow); e-mail: xvatkoff@mail.ru; Moscow, Russia; phone: 89856834525; deputy chief.

Khvatkov Vadim Aleksandrovich – The military training center at the National research University "Moscow Institute of electronic technology" (Moscow); xvatkoff@mail.ru; Moscow, Zelenograd, Russia, phone: +79264608261; head of the educational department – deputy head.

УДК 681.51-355.11

DOI 10.18522/2311-3103-2021-4-209-219

С.А. Боцвин, В.А. Хватков

ТРАНСФОРМАЦИЯ И АНАЛИЗ ИНФОРМАЦИИ ПРИ СОЗДАНИИ БАЗЫ ДАННЫХ УЧАСТНИКОВ ВЕЛИКОЙ ОТЕЧЕСТВЕННОЙ ВОЙНЫ 1941–1945 ГОДОВ МЕМОРИАЛЬНОГО КОМПЛЕКСА «ДОРОГА ПАМЯТИ» ГЛАВНОГО ХРАМА ВООРУЖЕННЫХ СИЛ РОССИЙСКОЙ ФЕДЕРАЦИИ НА ОСНОВЕ КОМПЬЮТЕРНЫХ МЕТОДОВ ОБРАБОТКИ ИНФОРМАЦИИ

Сохранение исторической памяти об участниках Великой Отечественной войны 1941–1945 годов является задачей мирового уровня, которая должна сберечь правду о самой страшной войне и подвиге нашего народа. В современных условиях привлечение интереса к истории, традициям и наконец признанию своего долга перед ушедшими поколениями требует современных методов. Одним из таких методов является трансформация информации, которая позволяет представить эту информацию в таком виде, чтобы она могла быть использована наиболее эффективно. При этом основной целью при трансформации исторических данных заключается в оптимизации их представлений и форматов и не изменении информационного содержания. Представленные алгоритмы трансформации и анализа информации при создании базы данных участников Великой Отечественной войны были направлены на максимальное сохранение исторической ценности и достоверности информации. Для достижения этой цели рассмотрены компьютерные методы обработки информации для нормализации и консолидации персональных данных, полученных из различных источников. Проведен анализ содержания сведений в архивных документах с представлением статистических данных по количеству документов (записей) из различных источников (архивов, базах данных, информационных ресурсов и т.п.) и описан порядок перевода информации из архивных документов в электронный вид, который был применен на практике. На основании анализа сведений построены диаграммы содержания персональных сведений в архивных источниках, представлены этапы систематизации и приведения к единому формату записей обобщенного информационного массива, а также порядок объединения и удаления дублирующих записей. Для возможности использования в других проектах подробно изложен алгоритм консолидации данных, полученных из различных источников, и построена его блок-схема. Кроме того описаны примененные алгоритмы нечеткого поиска, которые позволили минимизировать ошибки в записях, а также алгоритмы сравнения изображений для поиска дубликатов по фотографиям. Все приведенные алгоритмы позволили собрать воедино информацию, содержащуюся на различных носителях, имеющую разные структуры и географическое положение. Созданный информационный ресурс позволяет колоссально сократить ресурсы необходимые для

поиска нужных сведений, в том числе доступ к которым был ограничен или вообще отсутствовал. Дальнейшее усовершенствование алгоритмов нормализации и консолидации информации может послужить основой для миграции данных из устаревших в перспективные системы, а также для формирования информационных ресурсов из имеющихся разнородных архивных фондов.

Трансформация и анализ информации; учет личного состава; компьютерные методы обработки информации; алгоритмы нормализации и консолидации данных.

S.A. Botsvin, V.A. Khvatkov

TRANSFORMATION AND ANALYSIS OF INFORMATION WHEN CREATING A DATABASE OF PARTICIPANTS OF THE GREAT PATRIOTIC WAR 1941–1945 IN THE MEMORIAL COMPLEX «ROAD OF MEMORY» IN THE MAIN RUSSIAN ARMED FORCES CATHEDRAL ON THE BASIS OF COMPUTER METHODS OF INFORMATION PROCESSING

Preserving the historical memory of the participants of the Great Patriotic War 1941–1945 is a world-class task that should preserve the truth about the most terrible war and the feat of our people. In modern conditions, attracting interest in history, traditions and finally recognition of one's duty to the past generations requires modern methods. One of these methods is the transformation of information, which allows you to present this information in such a way that it can be used most effectively. At the same time, the main goal in the transformation of historical data is to optimize their representations and formats and not change the information content. The presented algorithms of transformation and analysis of information when creating a database of participants of the Great Patriotic War were aimed at maximizing the preservation of historical value and reliability of information. To achieve this goal, computer methods of information processing for normalization and consolidation of personal data obtained from various sources are considered. The analysis of the content of information in archival documents with the presentation of statistical data on the number of documents (records) from various sources (archives, databases, information resources, etc.) is carried out and the procedure for translating information from archival documents into electronic form, which has been applied in practice, is described. Based on the analysis of the information, diagrams of the content of personal information in archival sources are constructed, the stages of systematization and bringing the generalized information array records to a single format are presented, as well as the procedure for combining and deleting duplicate records. For the possibility of using in other projects, an algorithm for consolidating data obtained from various sources is described in detail, and its block diagram is constructed. In addition, the applied fuzzy search algorithms are described, which made it possible to minimize errors in records, as well as image comparison algorithms for searching for duplicates from photographs. All of these algorithms have made it possible to bring together information contained on various media, having different structures and geographical location. The created information resource allows you to enormously reduce the resources needed to find the necessary information, including access to which was limited or not at all. Further improvement of algorithms for normalization and consolidation of information can serve as a basis for data migration from outdated to promising systems, as well as for the formation of information resources from existing heterogeneous archival funds.

Transformation and analysis of information; personnel accounting; computer methods of information processing; algorithms for data normalization and consolidation.

Введение. Президентом Российской Федерации 2020 год был объявлен Годом памяти и славы для сохранения исторической памяти и в ознаменование 75-летия Победы в Великой Отечественной войне 1941–1945 годов. Завершено строительство Главного храма Вооруженных Сил Российской Федерации в историко-мемориальном комплексе которого создана «Дорога памяти» где в мозаичных картинах о войне используются фотографии и персональные данные из базы данных участников Великой Отечественной войны 1941–1945 годов (ВОВ).

Для формирования информационного массива участников ВОВ были привлечены Центральный архив и военно-учебные заведения Минобороны России, организовано взаимодействие с краеведческими (историческими) музеями, органами местного самоуправления, ветеранскими и иными организациями. Проведена работа с министерствами обороны иностранных государств и получены сведения из Казахстана, Сербии, Молдавии, Словакии, Беларуси и др. Осуществлен сбор фотографий и информации из семейных архивов населения через пункты отбора на военную службу по контракту и военные комиссариаты. Во время выездных информационно-агитационных акций «Военная служба по контракту в Вооруженных Силах Российской Федерации – Твой выбор!», пунктами отбора на военную службу по контракту велось дополнительное информирование населения о возможности представления персональных сведений на участников ВОВ с привлечением средств массовой информации. Организована работа пунктов сбора и загрузки сведений от населения, включая оригиналы фронтовых писем, в военно-учебных заведениях и довузовских образовательных организациях.

Актуальность. В целях получения достоверной информации об участниках ВОВ, проведена кропотливая работа по анализу более 230 миллионов архивных документов из различных источников. Изучены личные дела офицеров, послужные карты военнослужащих, учетные документы политработников, учетные документы военнопленных, военно-медицинские документы, список безвозвратных потерь личного состава Красной Армии, а также учтены сведения на 2,5 миллиона персоналий из информационных ресурсов глобальной информационной сети «Интернет», объединенных баз данных «Мемориал», «Подвиг народа» и «Память народа». Кроме того, в проведенной работе использовалась информация из печатных изданий, содержащих более 600 тысяч записей об участниках ВОВ. Дополнительно, сформированы информационные массивы данных о Героях СССР, кавалерах Ордена Славы и гражданах иностранных государств, награжденных Советскими орденами и медалями.

Оцифровка архивных документов и формирование электронных ресурсов обеспечивают открытость доступа к информации и сохранность подлинников, повышают эффективность и скорость поиска, увеличивают количество и географию пользователей (читателей), при этом работа с архивными документами ВОВ является важной государственной задачей мирового уровня, направленной на сохранение общей исторической памяти, правды о страшной войне и подвига воинов нашей Родины.

Компьютерные методы обработки информации при создании базы данных участников ВОВ. В ходе обработки полученной информации, применялись инновационные методы работы с большими массивами данных, такие как программное распознавание лиц, проверка на одну и две опечатки в персональных данных и алгоритм консолидации сведений из различных источников на одного участника ВОВ, позволившие исключить возможное дублирование или некорректный ввод информации.

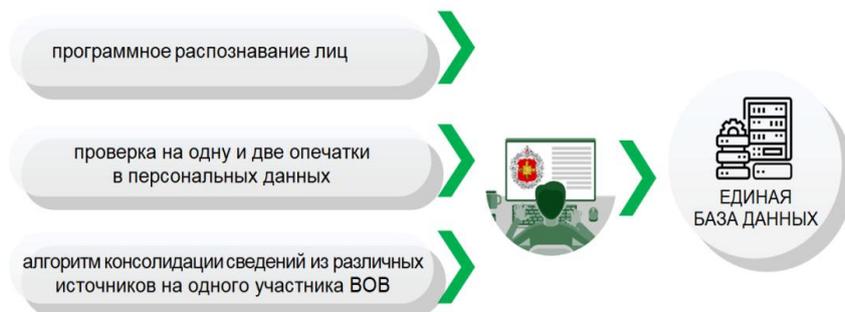


Рис. 1. Схема перевода информации из архивных документов в электронный вид

Перевод информации из архивных документов в электронный вид осуществлялся по следующему алгоритму:

- ◆ массовый перевод (сканирование) документов и картотек в электронный вид;
- ◆ систематизация отсканированных документов;
- ◆ ручная обработка персональных сведений;
- ◆ формирование классификаторов и справочников по воинским званиям, наградам, военным комиссариатам и местам рождениям, в том числе с учетом изменения наименований населённых пунктов и административных делений;
- ◆ заведение информации в базу данных;
- ◆ обобщение в единую электронную базу данных;
- ◆ систематизация записей и устранение дублей.

Источниками информации для формирования единой электронной базы данных явились (рис. 2): послужные картотеки и личные дела офицерского состава более 33 млн. документов; медицинская картотека ранений, которая заполнялась в различных лечебных заведениях на больных, раненых и умерших – 25 млн. карточек; базы данных порталов «Память народа», «Мемориал», «Бессмертный полк», «Подвиг народа» более 120 млн. записей; печатные издания с биографическими сведениями участников ВОВ – 600 тыс. записей; сведения из семейных архивов, представленные родственниками – 400 тыс. записей. При этом необходимо отметить, что все источники имели различную структуру и характер содержащейся информации.

Информационный массив приведен к единому формату по тринадцати полям: номер по порядку, воинское звание, фамилия, имя, отчество, дата рождения, место рождения, дата призыва, кем призван, дата гибели (смерти), сведения о награждении, фотография и примечание.



Рис. 2. Источники данных для формирования информационного массива

Систематизация записей участников ВОВ и приведение к единому формату осуществлялись в несколько этапов. На первоначальном этапе были определены наиболее приоритетные источники информации, то есть проведена экспертная оценка достоверности, полноты и качества источников информации и выполнена их сортировка. Основой формируемого единого массива стали сведения, полученные при сканировании и распознавании личных дел и послужных картотек участников ВОВ, хранящихся в архивных фондах Минобороны России. Далее созданный

информационный массив дополнялся недостающими сведениями из источников, имеющих более низкую приоритетность, а иногда и требующих дополнительного подтверждения.

Отдельного внимания заслуживает этап объединения записей и устранения дублирующих сведений. В связи с тем, что источники информации имели различную структуру и содержание, порядок внесения сведений, даты записей, а также содержали ошибки и неточности, были применены различные компьютерные методы обработки информации. При этом наличие ошибок было обусловлено некачественным распознаванием текстов с графических образов документов и ветхого состояния оригиналов, неточностями в учетных документах и, конечно же, «человеческим фактором» при ручной обработке информации. Также нельзя забывать, что учетные документы периода ВОВ имели широкую географию и источники формирования и хранения: человек мог быть призван в Сибири (именные списки призванных, формируемых военными комиссариатами), проходить службу на Западном фронте (послужная карта, заполняемая кадровым органом войсковой части), находиться на излечении в Крыму (карточка учета поступивших в лечебное учреждение), иметь наградные материалы (наградные листы, ходатайства и т.д.), подготовленные вышестоящим командованием. Кроме того, ведение послужных картотек в тот период было организовано только для офицерского состава, учет рядового состава осуществлялся по именным спискам и приказам командиров воинских частей.

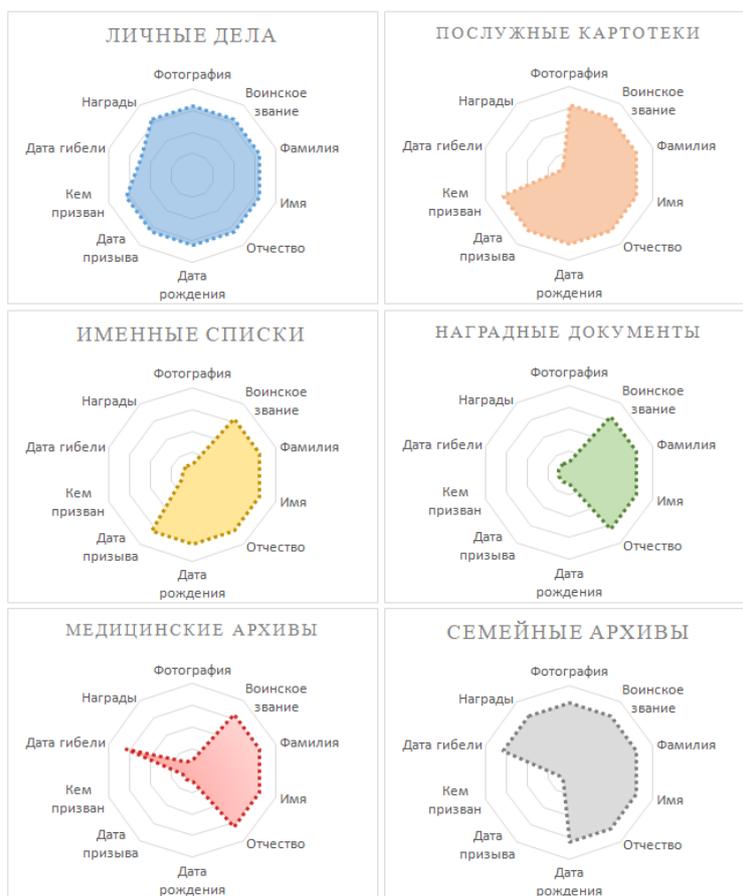


Рис. 3. Диаграммы содержания сведений в архивных источниках

Для приведения информационного массива к единому формату был проведен анализ содержания необходимых сведений по каждому классу архивных документов и составлены диаграммы в абсолютных величинах по содержанию сведений, которые представлены на рис. 3 в процентном соотношении. Необходимо отметить, что личные дела и послужные картотеки офицерского состава периода ВОВ позволили сформировать наиболее полные записи в едином массиве, а для рядового состава, чья численность значительно превышала численность офицеров – записи формировались путем объединения записей из различных источников в несколько этапов.

Алгоритм консолидации сведений из различных источников на одного участника ВОВ. После определения достоверности и приоритетности источников информации были проведены работы по формированию единого информационного массива по следующему алгоритму (рис. 4):

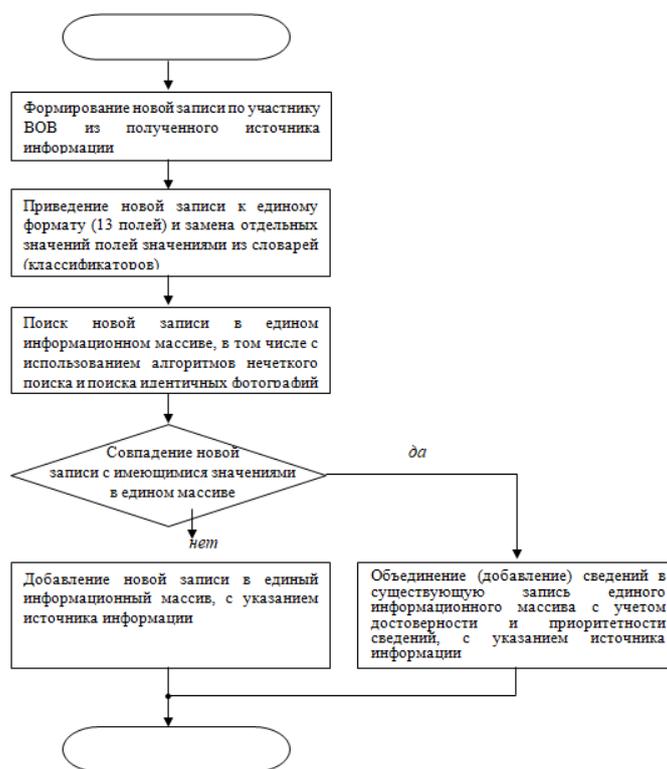


Рис. 4. Алгоритм консолидации сведений из различных источников

Использование такого алгоритма позволило организовать работу, не зависящую от сроков поступления информации, её полноты и объемов. Кроме того, этот алгоритм был неоднократно использован для поиска дублирующихся записей в едином массиве после внесения (добавления) сведений в существующие записи.

При выполнении поиска дублирующихся записей успешно были применены алгоритмы на частичное совпадение данных, то есть из подпрограмм сравнения записей попеременно исключались некоторые значения полей, а также использовались алгоритмы нечеткого поиска основанные на расстояниях Левенштейна и Дамерау – Левенштейна, которые позволили минимизировать ошибки в записях, причинами возникновения которых явились: «ручной ввод», некачественное распознавание

графических образов и наличие опечаток. Сравнение изображений (поиск дубликатов по фотографиям) осуществлялось путем приведения изображений к единому размеру (сжатию), «отбрасыванию» цветовой информации (преобразование изображения), нахождения среднеквадратической разности для каждой пары преобразованных изображений и сравнение полученной среднеквадратической разности с некоторым порогом и дополнительного обнаружения контура лица на изображении методом Виолы-Джонса и определения ключевых точек при помощи подхода, основанного на решающих обучаемых деревьях. Алгоритм поиска дубликатов по фотографиям показал высокую эффективность для людей, сменивших фамилию, имя или отчество, и особенно актуально этот вопрос был затронут при объединении записей для военнослужащих женского пола.

Выводы. Оцифровка архивных документов и формирование единых информационных ресурсов необходимы для обеспечения открытости доступа информации, уменьшения необходимого времени для её поиска, сохранения исторической памяти и увеличению количества и географии пользователей.

Работы выполненные при создании базы данных участников Великой Отечественной войны 1941–1945 годов мемориального комплекса «Дорога памяти» наглядно показывает, как с использованием компьютерных методов обработки информации возможно осуществить трансформацию и анализ информации. При этом используемая информация может содержаться на различных носителях, иметь разные структуры, объемы, географическое положение и для её поиска требуется огромное количество ресурсов, а доступ к сведениям может быть ограничен или вообще отсутствовать для обычного пользователя.

Предложенные алгоритмы обработки информации позволили в достаточно короткие сроки осуществить обработку колоссального количества сведений, содержащихся на различных носителях информации, её нормализацию и верификацию.

Дальнейшее направление исследования в этой области предлагается направить на усовершенствование алгоритмов нормализации и консолидации информации, полученной из различных источников, которые также могут послужить основой для разработки алгоритмов для миграции данных из устаревших систем в перспективные системы учета личного состава Вооруженных Сил Российской Федерации, а также при формировании информационных ресурсов, направленных на учет личного состава или создание из имеющихся фондов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. О Стратегии развития информационного общества в Российской Федерации на 2017-2030 годы: указ Президента РФ от 09.05.2017 № 203. – URL: <http://www.kremlin.ru/acts/bank/41919> (дата обращения: 05.04.2021).
2. О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года: указ Президента РФ от 07.05.2018 № 204. – URL: <http://www.kremlin.ru/acts/bank/43027> (дата обращения: 02.05.2021).
3. *Микрин Е.А. и др.* Информационное обеспечение систем организационного управления (теоретические основы): в 3-х ч. Ч. 3. Методы повышения качества информационного обеспечения систем организационного управления / под ред. Е.А. Микрина, В.В. Кульбы. – М.: Физматлит, 2012. – 528 с.
4. Правила организации, хранения, комплектования, учета и использования документов Архивного фонда Российской Федерации и других архивных документов в государственных и муниципальных архивах, музеях и библиотеках, организациях Российской академии наук. – М., 2007.
5. Рекомендации по комплектованию, учету и организации и хранению электронных архивных документов в государственных и муниципальных архивах. – М., 2014.
6. Правила издания исторических документов в СССР. Главархив СССР, ВНИИДАД, АН СССР, Ин-т истории СССР, Археогр. Комис. ИМЛ при ЦК КПСС, Гос. ком. СССР по печати, МГИАИ. – 2-е изд., перераб. и доп. – М., 1990. – 187 с.

7. Вебер Х., Дерр М. Оцифровка как метод обеспечения сохранения // Научные и технические библиотеки. – 1998. – № 10.
8. Веретехина С.В. Оцифровка архивных документов в формате PDF/A // Инновации в науке. – 2016. – № 2 (51). – URL: <https://cyberleninka.ru/article/n/otsifrovka-arhivnyh-dokumentov-v-formate-pdf-a> (дата обращения: 21.03.2021).
9. Греф Г. Информационные технологии - отстой: будущее за экономикой данных. – URL: <http://2035.media/2017/10/24/gref-future/> (дата обращения 16.05.2021).
10. Горемыкин В.П. Главное управление кадров Министерства обороны Российской Федерации: история и современность // Военная мысль. – 2013. – № 6. – С. 3-10.
11. Гусенков А.М. Интеллектуальный поиск сложных объектов в массивах больших данных // Российский научный электронный журнал. – 2016. – Т. 19, № 1. – С. 40-76.
12. Дюк В. Data Mining - интеллектуальный анализ данных. – Режим доступа: https://iteam.ru/publications/it/section_92/article_1448/ (дата обращения: 14.06.2021).
13. Ермолова В.В., Ермолова А.С., Ланская Д.В. Электронные документы, оцифровка и археология знаний // ЕГИ. – 2020. – № 3 (29). – URL: <https://cyberleninka.ru/article/n/elektronnye-dokumenty-otsifrovka-i-arheologiya-znaniy> (дата обращения: 21.10.2021).
14. Кульба В.В., Мамиконов А.Г., Пелихов В.П. Методы повышения достоверности и сохранности информации в АСУ: Обзор // Автоматика и телемеханика. – 1985. – № 2. – С. 5-33.
15. Лисицин А.Л. Программный комплекс оцифровки документов на бумажной основе // Computational nanotechnology. – 2015. – № 3. – URL: <https://cyberleninka.ru/article/n/programmnyy-kompleks-otsifrovki-dokumentov-na-bumazhnoy-osnove> (дата обращения: 21.10.2021).
16. Маннинг К., Рагхаван П., Шютце Х. Введение в информационный поиск. – М.: Вильямс, 2015. – 528 с.
17. Морозова Т.Ю. Базовые технологии автоматизации производства // Промышленные АСУ и контролеры. – 2013. – № 6. – С. 3-8.
18. Павельева Е.А. Использование фазовой информации в обработке изображений // Тр. международной конференции по компьютерной графике и зрению GraphiCon2016. – 2016. – С. 476-483.
19. Путин В.В. 75 лет Великой Победы: общая ответственность перед историей и будущим: официальный сайт Президента Российской Федерации // Новости, Выступления и стенограммы. 2020. – URL: <http://kremlin.ru/events/president/news/63527> (дата обращения: 02.04.2021).
20. Хворова И.Е. Процесс оцифровки документов для создания электронного архива // История и архивы. – 2017. – № 2. – С. 22-29.
21. Эрнст В. Архивация: архив как хранилище памяти и его инструментализация при национал-социализме // Новое литературное обозрение. – 2005. – № 74. – URL: <http://magazines.russ.ru/nlo/2005/74/ern9.html> (дата обращения: 22.04.2021).
22. Юмашева Ю.Ю. Методические рекомендации по электронному копированию архивных документов и управлению полученным информационным массивом. – М.: ВНИИДАД, 2012. – 217 с.
23. Юмашева Ю.Ю. Научное издание исторических документов в электронной среде: проблемы источниковедения и археологии // Историческая информатика. – 2017. – № 1 (19). – URL: <https://cyberleninka.ru/article/n/nauchnoe-izdanie-istoricheskikh-dokumentov-v-elektronnoy-srede-problemy-istochnikovedeniya-i-arheologii> (дата обращения: 21.05.2021).
24. ISO 12653-1:2007. Обработка изображений электронная. Тест-объект для черно-белого сканирования офисных документов.
25. ISO 29861:2009. Прикладные системы управления документами. Контроль качества сканирования цветных учрежденческих документов.
26. Aggarwal C.C. Data streams: models and algorithms. – Springer Science & Business Media, 2007. – 353 p. – DOI: 10.1007/978-0-387-47534-9.
27. Khan F.A., Ahmad A., Imran M., Alharbi M., Jan B. Efficient data access and performance improvement model for virtual data warehouse // Sustainable cities and society. – 2017. – Vol. 35. – P. 232-240. – DOI: 10.1016/j.scs.2017.08.003.

REFERENCES

1. O Strategii razvitiya informatsionnogo obshchestva v Rossiyskoy Federatsii na 2017-2030 gody: ukaz Prezidenta RF ot 09.05.2017 № 203 [Strategy of information society development in the Russian Federation for 2017-2030 years: the decree of the President of the Russian Federation dated 09.05.2017 No. 203]. Available at: <http://www.kremlin.ru/acts/bank/41919> (accessed 05 April 2021).
2. O natsional'nykh tselyakh i strategicheskikh zadachakh razvitiya Rossiyskoy Federatsii na period do 2024 goda: ukaz Prezidenta RF ot 07.05.2018 № 204 [On national goals and strategic objectives of the development of the Russian Federation for the period up to 2024: Decree of the President of the Russian Federation dated 07.05.2018 No. 204]. Available at: <http://www.kremlin.ru/acts/bank/43027> (accessed 02 May 2021).
3. *Mikrin E.A. i dr.* Informatsionnoe obespechenie sistem organizatsionnogo upravleniya (teoreticheskie osnovy): v 3-kh ch. Ch. 3. Metody povysheniya kachestva informatsionnogo obespecheniya sistem organizatsionnogo upravleniya [Information support of organizational management systems (theoretical foundations): in 3 part. Part 3. Methods of improving the quality of information support of organizational management systems] ed. by E.A. Mikrina, V.V. Kul'by. Moscow: Fizmatlit, 2012, 528 p.
4. Pravila organizatsii, khraneniya, komplektovaniya, ucheta i ispol'zovaniya dokumentov Arkhivnogo fonda Rossiyskoy Federatsii i drugikh arkhivnykh dokumentov v gosudarstvennykh i munitsipal'nykh arkhivakh, muzeyakh i bibliotekakh, organizatsiyakh Rossiyskoy akademii nauk [Rules of organization, storage, acquisition, accounting and use of documents of the Archival Fund of the Russian Federation and other archival documents in state and municipal archives, museums and libraries, organizations of the Russian Academy of Sciences]. Moscow, 2007.
5. Rekomendatsii po komplektovaniyu, uchetu i organizatsii i khraneniyu elektronnykh arkhivnykh dokumentov v gosudarstvennykh i munitsipal'nykh arkhivakh [Recommendations on the acquisition, accounting and organization and storage of electronic archival documents in state and municipal archives]. Moscow, 2014.
6. Pravila izdaniya istoricheskikh dokumentov v SSSR. Glavarkhiv SSSR, VNIIDAD, AN SSSR, In-istorii SSSR, Arkheogr. Komis. IML pri TSK KPSS, Gos. kom. SSSR po pechatu, MGIAI [Rules for publishing historical documents in the USSR. Glavarkhiv of the USSR, VNIIDAD, Academy of Sciences of the USSR, Institute of History of the USSR, Archeogr. Komis. IML at the Central Committee of the CPSU, State. com. USSR Press, MGIAI]. 2nd ed. Moscow, 1990, 187 p.
7. *Veber Kh., Derr M.* Otsifrovka kak metod obespecheniya sokhraneniya [Digitization as a method of ensuring preservation], *Nauchnye i tekhnicheskie biblioteki* [Scientific and technical libraries], 1998, No. 10.
8. *Veretekhina S.V.* Otsifrovka arkhivnykh dokumentov v formate PDF/A [Digitization of archival documents in PDF format/A], *Innovatsii v nauke* [Innovations in Science], 2016, No. 2 (51). Available at: <https://cyberleninka.ru/article/n/otsifrovka-arhivnyh-dokumentov-v-formate-pdf-a> (accessed 21 March 2021).
9. *Gref G.* Informatsionnye tekhnologii - otstoy: budushchee za ekonomikoy dannykh [Information technology sucks: the future belongs to the data economy]. Available at: <http://2035.media/2017/10/24/gref-future/> (accessed 16 May 2021).
10. *Goremykin V.P.* Glavnoe upravlenie kadrov Ministerstva oborony Rossiyskoy Federatsii: istoriya i sovremennost' [The Main Personnel Directorate of the Ministry of Defense of the Russian Federation: history and modernity], *Voennaya mysl'* [Military Thought], 2013, No. 6, pp. 3-10.
11. *Gusenkov A.M.* Intellektual'nyy poisk slozhnykh ob'ektov v massivakh bol'shikh dannykh [Intelligent search for complex objects in big data arrays], *Rossiyskiy nauchnyy elektronnyy zhurnal* [Russian Scientific Electronic Journal.], 2016, Vol. 19, No. 1, pp. 40-76.
12. *Dyuk V.* Data Mining - intellektual'nyy analiz dannykh [Data Mining - data mining]. Available at: https://iteam.ru/publications/it/section_92/article_1448/ (accessed 14 June 2021).
13. *Ermolenko V.V., Ermoshina A.S., Lanskaya D.V.* Elektronnye dokumenty, otsifrovka i arkhologiya znaniy [Documentos electrónicos, digitalización y arqueología del conocimiento], *EGI* [Investigación natural-humanitaria], 2020, No. 3 (29). Available at: <https://cyberleninka.ru/article/n/electronnye-dokumenty-otsifrovka-i-arheologiya-znaniy> (accessed 21 October 2021).

14. Kul'ba V.V., Mamikonov A.G., Pelikhov V.P. Metody povysheniya dostovernosti i sokhrannosti informatsii v ASU: Obzor [Métodos para aumentar la confiabilidad y la seguridad de la información en ACS: Revisión], *Avtomatika i telemekhanika* [Automatización y telemecánica], 1985, No. 2, pp. 5-33.
15. Lisitsin A.L. Programmnyy kompleks otsifrovki dokumentov na bumazhnoy osnove [Software package for digitizing paper-based documents], *Computational nanotechnology*, 2015, No. 3. Available at: <https://cyberleninka.ru/article/n/programmnyy-kompleks-otsifrovki-dokumentov-na-bumazhnoy-osnove> (accessed 21 October 2021).
16. Manning K., Raghavan P., Shyuttse Kh. Vvedenie v informatsionnyy poisk [Introduction to information search]. Moscow: Vil'yams, 2015, 528 p.
17. Morozova T.Yu. Bazovye tekhnologii avtomatizatsii proizvodstva [Basic technologies of production automation], *Promyshlennye ASU i kontrolery* [Industrial automated control systems and controllers], 2013, No. 6, pp. 3-8.
18. Pavell'eva E.A. Ispol'zovanie fazovoy informatsii v obrabotke izobrazheniy [The use of phase information in image processing], *Tr. mezhdunarodnoy konferentsii po komp'yuternoy grafike i zreniyu GraphiCon2016* [Proceedings of the International Conference on Computer Graphics and Vision GraphiCon2016], 2016, pp. 476-483.
19. Putin V.V. 75 let Velikoy Pobedy: obshchaya otvetstvennost' pered istoriey i budushchim: ofitsial'nyy sayt Prezidenta Rossiyskoy Federatsii [75 years of the Great Victory: shared responsibility to history and the future: official website of the President of the Russian Federation], *Novosti, Vystupleniya i stenogrammy. 2020* [News, Speeches and transcripts. 2020]. Available at: <http://kremlin.ru/events/president/news/63527> (accessed 02 April 2021).
20. Khvorova I.E. Protssess otsifrovki dokumentov dlya sozdaniya elektronnoy arkhiva [The process of digitizing documents to create an electronic archive], *Istoriya i arkhivy* [History and archives], 2017, No. 2, pp. 22-29.
21. Ernst V. Arkhivatsiya: arkhiv kak khranilishche pamyati i ego instrumentalizatsiya pri natsional-sotsializme [Archiving: archive as a storage of memory and its instrumentalization under National Socialism], *Novoe literaturnoe obozrenie* [New Literary Review], 2005, No. 74. Available at: <http://magazines.russ.ru/nlo/2005/74/ern9.html> (accessed 22 April 2021).
22. Yumasheva Yu.Yu. Metodicheskie rekomendatsii po elektronnomu kopirovaniyu arkhivnykh dokumentov i upravleniyu poluchennym informatsionnym massivom [Methodological recommendations on electronic copying of archival documents and management of the received information array]. Moscow: VNIIDAD. 2012, 217 p.
23. Yumasheva Yu.Yu. Nauchnoe izdanie istoricheskikh dokumentov v elektronnoy srede: problemy istochnikovedeniya i arkhografii [Scientific publication of historical documents in the electronic environment: problems of source studies and archeography], *Istoricheskaya informatika* [Historical Informatics], 2017, No. 1 (19). Available at: <https://cyberleninka.ru/article/n/nauchnoe-izdanie-istoricheskikh-dokumentov-v-elektronnoy-srede-problemy-istochnikovedeniya-i-arheografii> (accessed 21 May 2021).
24. ISO 12653-1:2007. Obrabotka izobrazheniy elektronnyaya. Test-ob"ekt dlya cherno-belogo skanirovaniya ofisnykh dokumentov [ISO 12653-1:2007. Image processing is electronic. Test object for black-and-white scanning of office documents].
25. ISO 29861:2009. Prikladnye sistemy upravleniya dokumentami. Kontrol' kachestva skanirovaniya tsvetnykh uchrezhdencheskikh dokumentov [ISO 29861:2009. Applied document management systems. Quality control of scanning of color institutional documents].
26. Aggarwal C.C. Data streams: models and algorithms. Springer Science & Business Media, 2007, 353 p. DOI: 10.1007/978-0-387-47534-9.
27. Khan F.A., Ahmad A., Imran M., Alharbi M., Jan B. Efficient data access and performance improvement model for virtual data warehouse, *Sustainable cities and society*, 2017, Vol. 35, pp. 232-240. DOI: 10.1016/j.scs.2017.08.003.

Статью рекомендовал к опубликованию д.т.н. А.В. Суворов.

Ботвин Сергей Александрович – Главное управление кадров Министерства обороны Российской Федерации (г. Москва); e-mail: xvatkoff@mail.ru; г. Москва, Россия, тел.: 89856834525; зам. начальника.

Хватков Вадим Александрович – Военный учебный центр при Национальном исследовательском университете «Московский институт электронной техники» (г. Москва); e-mail: xvatkoff@mail.ru; г. Москва, г. Зеленоград, Россия; тел.: 89264608261; начальник учебной части – заместитель начальника.

Botsvin Sergey Aleksandrovich – The Main personnel Directorate of the Ministry of defense of the Russian Federation (Moscow); e-mail: xvatkoff@mail.ru; Moscow, Russia; phone: 89856834525; deputy chief.

Khvatkov Vadim Aleksandrovich – The military training center at the National research University "Moscow Institute of electronic technology" (Moscow); Moscow, Zelenograd, Russia, phone: +79264608261; head of the educational department – deputy head.