

Проблема определения электронного документа долговременного хранения

А. В. Соловьев

Федеральный исследовательский центр "Информатика и управление" РАН, г. Москва, Россия

Аннотация. В статье исследованы определения понятий электронный документ, существующие в отечественных и международных нормативных документах. На основании проведенного исследования определены важные составные части электронного документа долговременного хранения: метаданные, нормализованная копия, связанные справочники и классификаторы, индексы, компоненты электронной подписи и др. Выполнено определение электронного документа в применении к организации долговременного хранения и определен, в общем виде, состав его информации, цифровых данных. Представлена математическая модель электронного документа в общем виде. Определены пути дальнейших научных изысканий: разработка математической модели семантических блоков электронного документа; оценка отчуждаемости электронного документа от программно-аппаратной среды хранения; оценка интерпретируемости электронных документов. В ходе дальнейших исследований автор планирует подготовить серию статей для описания решения проблем обеспечения сохранности документа.

Ключевые слова: долговременное хранение, сохранность документа, электронный документ, метаданные, цифровые данные.

DOI 10.14357/20718632220106

Введение

При разработке технологии обеспечения долговременного хранения цифровых данных важно комплексно решить проблему сохранности, обеспечив аутентичность, интерпретируемость, надежность и безопасность хранения [1]. Однако когда речь идет о долговременном хранении электронных документов (далее – ЭД), заменяющих бумажный документ в делопроизводстве и дальнейшем долговременном хранении в архиве, крайне важной становится проблема определения состава информации, необходимой и достаточной для полной интерпретации такого документа в будущем.

Деловой документ редко когда существует сам по себе. Как правило, он ссылается на другие документы, различные классификаторы

и справочные данные, нормативные документы и законодательные акты, базы данных и гиперссылки на интернет-ресурсы. Причем, классификаторы, справочные данные, нормативные документы могут со временем изменяться и дополняться, утрачивать силу и выпускаться в новой редакции. Тем самым, необходимо сохранять данные о классификаторах, справочниках, нормативных и иных документах в той их версии, которая существовала на момент поступления документа в долговременное хранение. Иначе семантика документа может быть со временем утрачена. Кроме того, документ может содержать в себе макросы, исправления и поля данных, которые могут привести к изменению документа в любой момент времени. Тогда можно утверждать, что, прежде чем приступить к организации долговременного хране-

ния электронных документов, необходимо ответить на следующие вопросы:

- что такое электронный документ?
- каков должен быть состав информации электронного документа?

1. Сложность определения понятия электронный документ

То, что определить понятие электронный документ не так просто, говорит наличие множества самых разных определений понятия электронный документ. Многие определения закреплены в нормативных документах. Рассмотрим наиболее актуальные из них.

Федеральный закон 2006 г. [2] определяет электронный документ следующим образом:

«Электронный документ – документированная информация, представленная в электронной форме, то есть в виде, пригодном для восприятия человеком с использованием электронных вычислительных машин, а также для передачи по информационно-телекоммуникационным сетям или обработки в информационных системах» [2, Ст. 2. п. 11.1]. Где «документированная информация – зафиксированная на материальном носителе путем документирования информация с реквизитами, позволяющими определить такую информацию, или, в установленных законодательством Российской Федерации случаях, ее материальный носитель» [2, Ст. 2. п. 11]. Здесь «информация — сведения (сообщения, данные) независимо от формы их представления» [2, Ст. 2. п. 1]. Это определение можно считать основным определением ЭД в РФ в настоящее время: оно наиболее полно и подробно описывает ЭД. Однако оно не определяет ЭД с точки зрения состава информации необходимой и достаточной для долговременного хранения.

В Федеральном законе 2011 г. [3] дается более общее определение ЭД. «Электронный документ – форма подготовки, отправления, получения или хранения информации с помощью электронных технических средств, зафиксированная на магнитном диске, магнитной ленте, лазерном диске и ином электронном материальном носителе» [3, Глава 1. Статья 2]. В данном определении, как и в предыдущем, ничего не говорится о составе информации для органи-

зации долговременного хранения. ЭД в этих определениях четко привязан к электронному материальному носителю информации. Это ограничение автоматически не рассматривает ЭД передаваемый по сети как документ.

Интересными с точки зрения состава информации ЭД является определение из [4]. Так, согласно этому стандарту «конструкторский документ в электронной форме (электронный документ) – документ, выполненный как структурированный набор данных, создаваемых программно-техническим средством» [4, ст. А.3]. В [4] дано также определение мультимедийного электронного документа «аудиовизуальный документ (мультимедийный документ) - электронный документ, содержащий видео- и (или) звуковую информацию» [4, ст. А.6]. В определениях важно, что в понятие ЭД включено понятие структуры данных документа. Правда, сама структура не определена. Также документ не привязан к электронному носителю информации. Но при этом ЭД как бы повисает в воздухе и получается оторванным от электронных средств и сред.

В стандарте [5] дано следующее, очень подробное определение ЭД. «Электронный документ – форма представления документа в виде множества взаимосвязанных реализаций в электронной среде и соответствующих им взаимосвязанных реализаций в цифровой среде» ([5], ст. 4, п.4.2.12). В стандарте даны определения реализации ЭД и воспроизведение ЭД. «Реализация электронного документа – отдельный элемент множества, представляющего электронный документ, существующая или могущая существовать в части электронной или цифровой среды» [5, ст. 4, п.4.3.1]. «Воспроизведение электронного документа – реализация электронного документа, доступная непосредственному восприятию человеком» [5, ст. 4, п.4.3.4]. В этом же стандарте дано следующее определение документа: «Документ – объект информационного взаимодействия в социальной среде, предназначенный для формального выражения социальных отношений между другими объектами этой среды» [5, ст. 4, п.4.1.1]. Кроме того, определены понятия электронная среда и цифровая среда. «Электронная среда – среда технических устройств (аппаратных

средств), функционирующих на основе физических законов и используемых в информационной технологии при обработке, хранении и передаче данных» [5, ст. 4, п.4.2.10]. «Цифровая среда – среда логических объектов, используемая для описания (моделирования) других сред (в частности, электронной и социальной) на основе математических законов» [5, ст. 4, п.4.2.11]. Т.е., согласно этим определениям ЭлД уже не привязан только к электронному носителю (электронной среде), а может оставаться документом в цифровой среде как логический объект (или множество логических объектов) или реализация документа, связанная с документом на электронном носителе. Кроме того, введены важные понятия реализации ЭлД при его отображении или передаче, а также воспроизведение – как отображение в памяти компьютера, на экране монитора для восприятия ЭлД пользователем.

В стандарте [6] также дано очень важное, на мой взгляд, определение ЭлД, вводящее важные реквизитные данные для идентификации ЭлД. «Электронный документ – информационный объект, состоящий из двух частей: реквизитной, содержащей идентифицирующие атрибуты (имя, время и место создания, данные об авторе и т. д.), электронную подпись, и содержательной, включающей в себя текстовую, числовую и/или графическую информацию, которая обрабатывается в качестве единого целого» [6], глава 3, п.3.2.22]. «Информационный объект – совокупность данных и программного кода, обладающая свойствами (атрибутами) и методами, позволяющими определенным образом обрабатывать данные ... самостоятельная единица применения и хранения в интегрированной информационной среде» [6, глава 3, п.3.2.2]. Интегрированная информационная среда (ИИС) – «совокупность распределенных баз данных, содержащих сведения об изделиях, производственной среде, ресурсах и процессах предприятия, обеспечивающая корректность, актуальность, сохранность и доступность данных тем субъектам производственно-хозяйственной деятельности (ПХД), участвующим в осуществлении жизненного цикла изделия (ЖЦИ), кому это необходимо и разрешено» [6, глава 3, п.3.2.1]. Важным в данном определении является деление ЭлД на части. Хотя такое деление вызывает вопросы: на

пример, зачем обязательно включать электронную подпись, зачем приравнивать ЭлД к информационному объекту. Но тем не менее, важно отметить, что ЭлД воспринимается как самостоятельная единица хранения, может включать в себя не только цифровые данные. Кроме всего прочего, ЭлД связывается с информационной средой и жизненным циклом некоторого материального объекта (изделия) реального мира. Правда информационная среда понимается исключительно как совокупность распределенных баз данных.

Интересным с точки зрения состава информации является определение, приведенное в [7]. «Электронный документ – документ на цифровом носителе, состоящий из трех элементов: содержание информации; форма предоставления содержания; носитель информации». Что важно, в этом определении вводится разделение содержания ЭлД от формы представления содержания. Однако документ строго привязывается к цифровому носителю информации как обязательной части документа, а не электронной среды, как в определении выше. Интересно, что стандарт ГОСТ Р 54989-2012 /ISO TR 18492:2005 «Обеспечение долговременной сохранности электронных документов» не содержит определения ЭлД, также отсутствует понятие о характеристиках ЭлД.

Теперь рассмотрим определения ЭлД существующие в мире и отличающиеся от приведенных выше определений. Определению ЭлД в иностранной литературе могут соответствовать три определения: *electronic document (e-document)*, *electronic record* и *digital document*.

В [8-10] ЭлД определен следующим образом. «An electronic document is any electronic media content (other than computer programs or system files) that are intended to be used in either an electronic form or as printed output». Т.е. электронным документом является любое содержимое (кроме компьютерных программ или системных файлов) электронных носителей, которое предназначено для использования в любом электронном виде или может быть представлено в виде печатной продукции. Согласно этому определению ЭлД – это любая последовательность бит, привязанная к носителям информации или каналам связи, которую пользователь может как-то увидеть или рас-

печатать. Причем, к ЭлД не относятся программы (исполняемые модули) и системные файлы операционных систем и других программных средств.

Термин electronic record, как правило, в литературе привязан к целому классу информационных систем, называемых системами электронного документооборота (СЭД), и является, по сути, их продуктом. В иностранной литературе термину СЭД соответствует Electronic Record Management System (ERMS), т.е. система управления служебными (деловыми) электронными документами. «Electronic record – служебный документ, который существует в электронной форме. Документ может оказаться представленным в электронной форме в результате того, что он изначально создается при помощи прикладной программы или в результате оцифровки, т.е. путем сканирования бумажного документа или микрофильма» [11]. В приведенном определении ЭлД – это и созданный программными средствами деловой документ, и оцифрованная копия бумажного документа. Отсутствует деление документа на части, не определен состав информации. ЭлД получается не связанным с электронным носителем.

Международный стандарт ISO 15489 Records Management в дополнение к определению ЭлД из [11] говорит о том, что «документ может состоять из одного или нескольких объектов данных», однако не определяет состав данных ЭлД. [11], являющийся фактическим стандартом для СЭД, также в разделе 10.3 вводит важное дополнение к определению ЭлД, а именно, что служебный документ должен быть защищен от изменений, а также быть привязан к своему жизненному циклу и системе классификации документов. В [11] вводится понятие метаданных, т.е. данных о ЭлД. Метаданные – «структурированная или полуструктурированная информация, которая дает возможность создания, управления, и использования документов в течение времени и внутри и вне организации внутри и вне области их создания». При этом признается, что «различие между данными и метаданными может быть неочевидно. Например, обычно понятно, что такие существенные индексные данные документа как заголовки, дата и т.д. есть часть метаданных документа. Однако данные системного журнала, относящиеся к документу, или порядок хранения

документа вполне могут рассматриваться как данные или как метаданные в зависимости от контекста. Различные типы метаданных могут быть определены, например, для индексирования, длительного хранения, отображения документа и т.д.». Впрочем, вопросы состава метаданных и конкретные рекомендации по обеспечению долговременной сохранности (отчасти потому, что МЭД для этого не предназначен) в [1] не рассматриваются. Однако разработчикам СЭД ставится задача, по крайней мере, продумать эти вопросы.

Термин digital document обычно уточняет понятие ЭлД: это документ, созданный по цифровой технологии, т.е. нечто, что можно представить в виде последовательности бит [12]. Можно привести многочисленные примеры из нормативных документов Германии, Дании, Франции, однако в этом нет большого смысла, так как они с той или иной степенью вариации повторяют вышеперечисленные определения.

Отсутствие единого определения термина ЭлД при организации долговременного хранения порождает целый ряд вопросов:

- как нужно определить ЭлД для организации долговременного хранения?
- нужно ли считать средства интерпретации и отображения ЭлД частью ЭлД?
- нужно ли кроме самого ЭлД хранить некоторую дополнительную информацию о нем?
- нужно ли разделять форму и содержание ЭлД?
- нужно ли разделять форму хранения и форму отображения ЭлД?
- каким должен быть состав информации для организации долговременного хранения ЭлД?

Поиск ответов на данные вопросы порождает необходимость решения частной научно-технической задачи, а именно, создание математической модели ЭлД при долгосрочном хранении, определяющей состав необходимой и достаточной информации для хранения.

2. Определение электронного документа долговременного хранения

В ИСА РАН (ФИЦ ИУ РАН) под руководством доктора технических наук, профессора

Н. Е. Емельянова была разработана теория документного интерфейса [13, 14], согласно которой документ (деловой документ) – это структурированная информация, представляющая собой совокупность взаимосвязанных семантических блоков. Семантические блоки — части документа, выделенные по смысловому содержанию. Действительно, всякий реальный деловой документ представляет собой совокупность взаимосвязанных частей (главы, разделы, подразделы, пункты и т.д.), которые и есть семантические блоки, т.к. каждый из них представляет собой законченный фрагмент документа с собственной семантикой (смысловым содержанием). Разумеется, деление документа на семантические блоки в общем случае неоднозначно. Необходимо разработать математическую модель документа и его семантических блоков в контексте долговременного хранения. Исходя из общего определения документа, можно определить электронный документ как документ, семантические блоки которого и взаимосвязи между ними представлены в электронно-цифровой форме.

В общем случае математическую модель ЭлД можно представить в виде графа (или дерева, если до корня из любой листовой вершины имеется единственный путь), вершинами которого являются семантические блоки B_i , ребрами – связи между B_i . B_i в свою очередь могут представлять собой подграфы (поддеревья), также состоящие из семантических блоков следующего уровня, т.к. в любом семантическом блоке почти всегда можно выделить заголовок, подзаголовки, повторяющиеся части (таблицы), агрегаты (массивы, структуры данных), атомарные цифровые данные и т.д. Все эти блоки также имеют свою оригинальную семантику. Деление документа на семантические блоки может осуществляться человеком, а может с помощью автоматической информационной системы [15].

Между различными ЭлД также могут существовать связи [16]. Тогда, множество связанных ЭлД, каждый из которых представляет собой отдельный граф из B_i , может быть объединено в единый граф связанных ЭлД. Учитывая рекомендации, необходимые для решения проблем, приведенных в [1], необходимо

отметить, что ЭлД, безусловно, должен содержать дополнительную информацию: метаданные, связь с классификаторами и справочниками, индексы (в том числе и полнотекстовые), ЭП (электронная подпись), сертификаты ЭП и СОС (списки отзыва сертификатов). Кроме этого, возможно включение в ЭлД данных журналов работы с ЭлД. Подобное разбиение ЭлД на семантические блоки или составные части согласуется с определениями ЭлД, подразумевающими, что ЭлД не просто информация, а набор необходимых для полной интерпретации ЭлД цифровых данных.

При долговременном хранении ЭлД кроме классификаторов и индексов, являющихся неотъемлемой частью ЭлД и проходящих вместе с ним возможные миграции в электронной и цифровой средах, ЭлД дополняется нормализованной копией документа. Нормализованная копия ЭлД необходима для повышения вероятности интерпретации через годы и десятилетия. Нормализованная копия представляет собой копию ЭлД, преобразованную в один из форматов долгосрочного хранения (открытых, документированных форматов) XML, ODF, PDF/A. Нормализованная копия может быть также представлена сочетанием форматов, например XML для хранения содержимого (текста) документа, метаданных, индексов, информацией о связи с другими документами. Если необходимо сохранить внешний вид документа, как можно более точно повторяющий внешний вид оригинала, то лучше использовать TIFF GROUP 4 для черно-белых документов или PNG для цветных (подробнее об этом - [17]).

Как видно из работы автора [1], ЭлД при долговременном хранении представляет собой объект управления в нестабильной программно-аппаратной среде, подверженной параметрическим возмущениям. Поэтому ЭлД должен быть максимально оторван от программно-аппаратной среды хранения (полный отрыв невозможен), что делает возможным его относительно безболезненный перенос (миграцию) из одной программно-аппаратной среды хранения в другую. Такая постановка задачи требует дополнительных исследований и разработки математической модели оценки отчуждаемости и интерпретируемости.

Согласно определению Элд данному выше в этом разделе, можно утверждать, что математическая модель Элд при долговременном хранении представляет собой объединение (оператор U) семантических блоков документа. В общем виде модель Элд определяется следующим образом: есть множество Элд $D = \{ DAr_i \}$, при этом: $DAr_i = U_{(i=1,N)}(B_i)$. Элд разбивается на семантические блоки, если:

- $B_i \cap B_j \neq \emptyset$, то $B_i \subseteq B_j$ или $B_j \subseteq B_i$;
- $\exists B_j \subset B_i$, то $B_i = \cup B_j$ для всех $B_j \subset B_i$.

Каждому Элд DAr_i из множества D поставим в соответствие некоторый граф $G_i(D) = (V, E)$, где V — множество вершин графа, E — множество дуг. $V = \{B_i\}$. $E = \{(B_i, B_j): B_i \rightarrow B_j\}$.

Каждый B_i представляет собой определенную часть Элд, фрагмент части (подчасть), фрагмент подчасти и т.д.

Тем самым, можно утверждать, что математическая модель Элд в общем виде определена. Необходимо определить необходимый состав семантических блоков B_i Элд и разработать математические модели для B_i .

Заключение

В результате проведенного исследования были исследованы определения Элд, существующие в отечественных и международных нормативных документах. Были определены важные составные части Элд для долговременного хранения, такие как: метаданные, нормализованная копия, связанные справочники и классификаторы, индексы, компоненты ЭП и др. Дано определение Элд в применении к организации долговременного хранения и определен в общем виде состав его информации. В общем виде разработана математическая модель Элд. Важность и актуальность данного исследования подчеркивается большим количеством работ, посвященных определению Элд и составу его информации. Это подтверждается приведенным кратким обзором в данной статье.

В проведенном исследовании определены пути дальнейших научных изысканий, а именно: разработка математической модели семантических блоков Элд долговременного хранения; разработка математической модели оценки отчуждаемости Элд от программно-аппаратной

среды хранения; разработка математической модели оценки интерпретируемости Элд.

В ходе дальнейших исследований автор планирует подготовить серию статей, где будут описаны пути решения перечисленных выше проблем. Это послужит еще одним шагом в создании технологии организации долговременного хранения Элд.

Литература

1. Solovyev, A.V. Long-Term Digital Documents Storage Technology // *Lecture Notes in Electrical Engineering*. 2020. vol.641, pp.901-911. ISSN 1876-1100. doi: 10.1007/978-3-030-39225-3_97.
2. Федеральный закон Российской Федерации от 27 июля 2006 года №149-ФЗ «Об информации, информационных технологиях и о защите информации».
3. Федеральный закон Российской Федерации от 6 апреля 2011 г. №63-ФЗ «Об электронной подписи».
4. ГОСТ 2.001-93 Единая система конструкторской документации. Общие положения.
5. ГОСТ Р 52292-2004 Информационная технология. Электронный обмен информацией. Термины и определения.
6. ГОСТ Р 50.1.031–2001 Рекомендации по стандартизации. Информационные технологии поддержки жизненного цикла продукции. Терминологический словарь. Часть 1. Стадии жизненного цикла продукции.
7. Большой юридический словарь / [А. В. Малько и др.]; под ред. А. В. Малько. - Москва: Проспект, 2009. 702 с.; ISBN 978-5-392-00675-5.
8. Biagi, S. Media/Impact. An Introduction to Mass Media // Wadsworth Publishing, 7th edition – 2004. 416 p.
9. Caron, A.H., Caronia, L. Moving cultures: mobile communication in everyday life // McGill-Queen's University Press. 2007. 280 p. ISBN 978-0-7735-3203-4.
10. McQuail, D. McQuail's Mass Communication Theory (fourth edition), Sage Publications Ltd., London – MAS. 2000. pp.16–34.
11. Typical requirements for automated electronic document management systems. Specification MoReq // Office for Official Publications of the European Communities as INSAR Supplement VI, ISBN 92-894-1290-9.
12. Buckland, M. What is a "digital document"? // *Document Numérique (Paris) 2* - №2. 1998. pp.221-230.
13. Емельянов, Н.Е. Виды представления структурированных данных // *Теоретические основы информационной технологии. Сборник трудов ВНИИСИ*. № 22, М.:ВНИИСИ, 1988. С.42–46.
14. Емельянов, Н.Е. Теоретический анализ документного интерфейса: Препринт / Н.Е. Емельянов. М.: Всесоюзный научно-исследовательский институт системных исследований. 1987, 40 с.
15. Grigoriev, O.G. et al. TITANIS: A Tool for Intelligent Text Analysis in Social Media // In: Kovalev S.M., Kuznetsov S.O., Panov A.I. (eds) *Artificial Intelligence. RCAI 2021. Lecture Notes in Computer Science*, vol

12948. pp. 232-247. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-86855-0_16.
16. Белова, А.Н., Соловьев, А.В. Построение баз данных взаимосвязанных документов // Труды ИСА РАН. 2012. 62(3). С.25-30.
17. Соловьев, А.В. Решение проблемы интерпретации цифровых данных долговременного хранения // Труды ИСА РАН. Том 71, выпуск 2, М.: 2021. С.43-49. DOI: 10.14357/20790279210206.

Соловьев Александр Владимирович. Федеральный исследовательский центр "Информатика и управление" РАН, г. Москва. Главный научный сотрудник, доктор технических наук. Количество печатных работ: 130. Область научных интересов: системный анализ, системы управления базами данных, теория надежности, математическое моделирование, долговременное хранение электронных документов. E-mail: soloviev@isa.ru

The Problem of Defining an Electronic Document for Long-Term Storage

A. V. Solovyev

Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia

Abstract. In this article, the definitions of the concepts of an electronic document that exist in domestic and international regulatory documents were investigated. Based on the study, important components of an electronic document for long-term storage were identified: metadata, a normalized copy, related directories and classifiers, indices, components of an electronic signature, etc. An electronic document was defined in relation to the organization of long-term storage and the composition of its information was determined in general terms. , digital data. The mathematical model of the electronic document is presented in general form. The ways of further scientific research are determined: development of a mathematical model of the semantic blocks of an electronic document; assessment of the alienability of an electronic document from the software and hardware storage environment; assessment of the interpretability of electronic documents. In the course of further research, the author plans to prepare a series of articles to describe the solution to the problems of ensuring the preservation of the document.

Keywords: long-term storage, document preservation, electronic document, metadata, digital data.

DOI 10.14357/20718632220106

References

1. Solovyev, A.V. 2020. Long-Term Digital Documents Storage Technology. Lecture Notes in Electrical Engineering. 641: 901-911. doi: 10.1007/978-3-030-39225-3_97.
2. Federal'nyy zakon Rossiyskoy Federatsii ot 27 iyulya 2006 goda №149-FZ «Ob informatsii, informatsionnykh tekhnologiyakh i o zashchite informatsii» [Federal Law of the Russian Federation dated July 27, 2006 No. 149-FZ "On Information, Information Technologies and Information Protection"]].
3. Federal'nyy zakon Rossiyskoy Federatsii ot 6 aprelya 2011 g. №63-FZ «Ob elektronnoy podpisi» [Federal Law of the Russian Federation of April 6, 2011 No. 63-FZ "On Electronic Signature"]].
4. GOST 2.001-93 Yedinaya sistema konstruktorskoy dokumentatsii. Obshchiye polozheniya [GOST 2.001-93 Unified system for design documentation. General provisions].
5. GOST R 52292-2004 Informatsionnaya tekhnologiya. Elektronnyy obmen informatsiyey. Terminy i opredeleniya [GOST R 52292-2004 Information technology. Electronic information exchange. Terms and Definitions].
6. GOST R 50.1.031–2001 Rekomendatsii po standartizatsii. Informatsionnyye tekhnologii podderzhki zhiznennogo tsikla produktsii. Terminologicheskii slovar'. Chast' 1. Stadii zhiznennogo tsikla produktsii [GOST R 50.1.031–2001 Recommendations for standardization. Information technologies for product life cycle support. Terminological dictionary. Part 1. Stages of the product life cycle].
7. A.V. Mal'ko et al. 2009. Bol'shoy yuridicheskii slovar' [Big Law Dictionary]. M: Prospekt. 702 p. ISBN 978-5-392-00675-5.
8. Biagi, S. 2004. Media/Impact. An Introduction to Mass Media. Wadsworth Publishing, 7th edition. 416 p.
9. Caron, A.H., Caronia, L. 2007. Moving cultures: mobile communication in everyday life. McGill-Queen's University Press. 280 p. ISBN 978-0-7735-3203-4.
10. McQuail, D. 2000. McQuail's Mass Communication Theory (fourth edition). Sage Publications Ltd., London – MAS. pp.16–34.
11. Typical requirements for automated electronic document management systems. Specification MoReq // Office for Official Publications of the European Communities as INSAR Supplement VI, ISBN 92-894-1290-9.

12. Buckland, M. 1998. What is a "digital document"? Document Numérique (Paris) 2. 2: 221-230.
13. Emelyanov, N.E. 1988. Vidy predstavleniya strukturovannykh dannykh [Types of representation of structured data]. Teoreticheskiye osnovy informatsionnoy tekhnologii. Sbornik trudov VNIISI. [Theoretical foundations of information technology. Collection of works of VNIISI]. 22: 42–46.
14. Emelyanov, N.E. 1987. Teoreticheskiy analiz dokumentnogo interfeysa [Theoretical analysis of the document interface]. Vsesoyuznyy nauchno-issledovatel'skiy institut sistemnykh issledovaniy [All-Union Research Institute for System Research]. 40 p.
15. Grigoriev, O.G. et al. 2021. TITANIS: A Tool for Intelligent Text Analysis in Social Media. Artificial Intelligence. RCAI 2021. Lecture Notes in Computer Science. Springer, Cham. 12948: 232-247. DOI: https://doi.org/10.1007/978-3-030-86855-0_16.
16. Belova, A.N., Solovyev, A.V. 2012. Postroyeniye baz dannykh vzaimosvyazannykh dokumentov [Building databases of related documents]. Trudy ISA RAN [Proceedings of the ISA RAS]. 62(3): 25-30.
17. Solovyev, A.V. 2021. Resheniye problemy interpretatsii tsifrovyykh dannykh dolgovremennogo khraneniya [Solving the problem of interpreting digital data for long-term keeping]. Trudy ISA RAN [Proceedings of the ISA RAS]. 71(2): 43-49. DOI: 10.14357/20790279210206.

Solovyev A. V. Chief Researcher, Doctor of Technical Sciences. Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, 44/2 Vavilova str., Moscow, 119333, Russia, e-mail: soloviev@isa.ru