

Математическое моделирование

Математическая модель отчуждаемости электронного документа от среды хранения

А.В. СОЛОВЬЕВ

Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия

Аннотация. В статье предложена математическая модель отчуждаемости электронного документа от программно-аппаратной среды хранения. Среда хранения электронного документа всегда подвижна и изменчива, подвержена параметрическим возмущениям. Следовательно, при долговременном хранении электронный документ должен быть, как объект управления, максимально стабилизирован по основной характеристике – сохранность. Для этого он, в частности, должен быть максимально независим от среды хранения. В качестве меры независимости предложена математическая модель оценки риска неотчуждаемости документа от среды хранения. Выполненные расчеты и практическое применение в ряде проектов электронного архива дают основания полагать адекватность представленной модели условиям задачи долговременного хранения.

Ключевые слова: *долговременное хранение, сохранность документа, электронный документ, отчуждаемость, оценка риска.*

DOI: 10.14357/20790279230401 **EDN:** LHCYGP

Введение

Как было показано в [1], при долговременном хранении электронных документов (далее – ЭЛД) крайне важной становится проблема определения состава информации необходимой и достаточной для полной интерпретации такого документа в будущем.

Согласно [2, 3] документ определен как структурированная информация, представляющая собой совокупность взаимосвязанных семантических блоков. Семантические блоки — части документа, выделенные по смысловому содержанию. Исходя из общего определения документа, можно определить электронный документ как документ, семантические блоки которого и взаимосвязи между ними представлены в электронно-цифровой форме.

В работе [4] определены семантические блоки для документа долговременного хранения и

предложена математическая модель долговременного хранения.

Однако важной нерешенной задачей остается определить насколько реально хранящийся ЭЛД, пусть созданный согласно правильной модели долговременного хранения, может быть сохранен при резком изменении программно-аппаратной среды хранения. Под резким изменением будем понимать существенные изменения, происходящие за непродолжительное время. Такими изменениями могут быть:

- выход из строя технических средств хранения (в первую очередь, носителей, на которых сохранены электронные документы);
- существенное обновление программных средств – операционные системы, программы интерпретации (чтения, декодирования и визуализация) электронных документов;

– изменение криптографических средств и стандартов защиты информации (если такие применяются при долговременном хранении).

Все эти изменения приводят к тому, что ЭЛД должны быть отчуждены (извлечены) от старой программно-аппаратной среды хранения и помещены в новую, где они также должны быть сохранены и интерпретированы.

Как показано в [5], программно-аппаратная среда хранения ЭЛД всегда подвижна и изменчива, подвержена параметрическим возмущениям. Следовательно, при долговременном хранении ЭЛД должен быть, как объект управления, максимально стабилизирован по основной характеристике – сохранность. Для этого он, в частности, должен быть максимально независим от среды хранения.

В качестве меры независимости в данной статье предложена математическая модель оценки риска неотчуждаемости ЭЛД от программно-аппаратной среды хранения.

Рассмотрим подробнее математическую модель отчуждаемости ЭЛД.

1. Математическая постановка задачи оценки отчуждаемости электронного документа

Для того, чтобы говорить о разработке математической модели оценки отчуждаемости (или риска неотчуждаемости, что в данном случае одно и то же), необходимо сформулировать постановку задачи.

Исходя из общей постановки задачи обеспечения долговременной сохранности, приведенной в [5], можно сформулировать задачу оценки отчуждаемости следующим образом.

Дано:

- 1) Множество ЭЛД $D = \{ D_k \}$;
- 2) Требования к допустимым значениям отчуждаемости φ_r ;
- 3) Математическая модель ЭЛД долговременного хранения D .

Найти: математическую модель оценки отчуждаемости $\varphi(t)$

Математическая модель ЭЛД долговременно хранения подробно рассмотрена в [4]. Приведем ее, т.к. она будет важна при решении поставленной задачи:

$$D_k = OrD \cup OdfD \cup DMD \cup CLI \cup LDI, \quad (1)$$

где OrD – электронный документ (оригинал) или оцифрованное изображение оригинального бумажного документа, которое далее также будем обозначать как оригинал;

DMD – данные о ЭЛД (метаданные), такие как автор (авторы), время и место создания, время последнего изменения, сведения о формате, название документа; возможны и другие данные, например, установленные стандартом «дублинского ядра» (Dublin Core) [6];

$OdfD$ – нормализованная копия ЭЛД; под нормализацией здесь понимается приведение ЭЛД к единому формату (набору форматов) долговременного хранения оригинала документа OrD ;

CLI – нормативно-справочная информация (классификаторы, словари, нормативные документы), на которые ссылается основной документ;

LDI – данные о документах, связанных с ЭЛД.

2. Математическая модель отчуждаемости электронного документа

Решением поставленной задачи должна быть математическая модель оценки отчуждения документа. Удобно представить оценку отчуждения как вероятность того, что ЭЛД может быть полностью извлечен из программно-аппаратной среды хранения и перенесен без потерь в новую среду. Под вероятностью в данной модели понимаем оценку риска неотчуждения, выполненную, например, с помощью экспертной оценки. При этом должны быть установлены уровни риска, например, с помощью вербальной шкалы из минимум 3-5 позиций (как это предложено, например, в [7]).

Такую оценку удобно представить как взвешенную свертку вероятностей отчуждения семантических блоков модели ЭЛД долговременного хранения (1).

Тогда математическая модель вероятности отчуждения ЭЛД может быть представлена в следующем виде:

$$\varphi_k(D_k) = k_1 p\varphi OrD + k_2 p\varphi OdfD + k_3 p\varphi DMD + k_4 p\varphi CLI + k_5 p\varphi LDI, \quad (2)$$

где k_i – весовые коэффициенты важности отчуждаемости семантических блоков ЭЛД (назначаются экспертами); $\sum k_i = 1, i=[1,5]$; если нет других предпочтений, то автор рекомендует назначать следующие значения $k_1 \geq 0,5-0,6, k_2 = \sum_{i=[3,5]} k_i$, т.к. отчуждаемость оригинала ЭЛД (OrD) и нормализованной копии ($OdfD$) является наиболее критичной при долговременном хранении ЭЛД;

$p\varphi OrD$ – вероятность отчуждения оригинала (OrD) ЭЛД (включая все его семантические блоки) от среды хранения;

$p\varphi OdfD$ – вероятность отчуждения нормализованной копии ($OdfD$) ЭЛД от среды хранения;

$p\phi DMD$ – вероятность отчуждения данных о Элд (DMD) от среды хранения;

$p\phi CLI$ – вероятность отчуждения, связанной с Элд нормативно-справочной информации (CLI) от среды хранения;

$p\phi LDI$ – вероятность отчуждения данных о связанных с Элд других документах (LDI) от среды хранения.

Разумеется, представленная математическая модель справедлива для оценки семантических блоков верхнего уровня. Каждый показатель формулы (2) необходимо детализировать, как детализована математическая модель Элд (см. [4]).

3. Математическая модель отчуждаемости оригинала электронного документа

В работе [4] приведена подробная математическая модель оригинала Элд, состав семантических блоков и их назначение. Если принять допущения, представленные в работе [4], математическая модель оценки отчуждаемости оригинала Элд типовой системы электронного документооборота (СЭД) представляется в следующем виде:

$$p\phi OrD = (\sum_{(i=1,N1)} kN1_i (p\phi OrDoc_i^{kN11} (\sum_{(j=1,M1)} kM1_{ij} p\phi DSign_{ij})^{kN12}))^{k1} (\sum_{(i=1,N2)} kN2_i (p\phi OrRes_i^{kN21} (\sum_{(j=1,M2)} kM2_{ij} p\phi RSign_{ij})^{kN22}))^{k2} (\sum_{(i=1,N3)} kN3_i (p\phi OrAgr_i^{kN31} (\sum_{(j=1,M3)} kM3_{ij} p\phi ASign_{ij})^{kN32}))^{k3} (\sum_{(i=1,N4)} kN4_i (p\phi OrExe_i^{kN41} (\sum_{(j=1,M4)} kM4_{ij} p\phi ESign_{ij})^{kN42}))^{k4} (\sum_{(i=1,N5)} kN5_i (p\phi OrMet_i^{kN51} (\sum_{(j=1,M5)} kM5_{ij} p\phi MSign_{ij})^{kN52}))^{k5} (\sum_{(i=1,N6)} kN6_i (p\phi OrApp_i^{kN61} (\sum_{(j=1,M6)} kM6_{ij} p\phi ApSign_{ij})^{kN62}))^{k6} p\phi SignOrD^{k7}, \quad (3)$$

где весовые коэффициенты важности, назначаемые экспертами или автоматически, в простейшем случае равные между собой, должны удовлетворять условиям:

$$\sum k_i = 1, i=[1,7],$$

$$\sum_{(i=1,N1)} kN1_i = 1, \sum_{(j=1,M1)} kM1_{ij} = 1, kN11 + kN12 = 1,$$

$$\sum_{(i=1,N2)} kN2_i = 1, \sum_{(j=1,M2)} kM2_{ij} = 1, kN21 + kN22 = 1,$$

$$\sum_{(i=1,N3)} kN3_i = 1, \sum_{(j=1,M3)} kM3_{ij} = 1, kN31 + kN32 = 1,$$

$$\sum_{(i=1,N4)} kN4_i = 1, \sum_{(j=1,M4)} kM4_{ij} = 1, kN41 + kN42 = 1,$$

$$\sum_{(i=1,N5)} kN5_i = 1, \sum_{(j=1,M5)} kM5_{ij} = 1, kN51 + kN52 = 1,$$

$$\sum_{(i=1,N6)} kN6_i = 1, \sum_{(j=1,M6)} kM6_{ij} = 1, kN61 + kN62 = 1,$$

где $p\phi OrDoc_i$ – вероятность отчуждения i -го семантического блока ($OrDoc_i$) оригинала Элд (например, если каждая страница многостраничного документа представлена отдельной оцифрованной

копией, несколько файлов составляют один документ (введение, разделы и заключение) и другие варианты деления), каждый из которых может быть заверен отдельным набором средств криптозащиты, например ЭП ($DSign_{ij}$);

$p\phi OrRes_i$ – вероятность отчуждения i -го оригинала листов резолюций Элд;

$p\phi OrAgr_i$ – вероятность отчуждения i -го оригинала листов согласования Элд;

$p\phi OrExe_i$ – вероятность отчуждения i -го оригинала листов исполнения Элд;

$p\phi OrMet_i$ – вероятность отчуждения i -го оригинала листов ознакомления Элд;

$p\phi OrApp_i$ – вероятность отчуждения i -го приложения Элд (например, файлы, в том числе аудио- и видео, изображения и др.);

$p\phi DSign_{ij}, p\phi RSign_{ij}, p\phi ASign_{ij}, p\phi ESign_{ij}, p\phi MSign_{ij}, p\phi ApSign_{ij}$ – вероятность отчуждения j -го компонента криптозащиты (ЭП) i -го семантического блока оригинала Элд, соответственно, $OrDoc_p, OrRes_p, OrAgr_p, OrExe_p, OrMet_p, OrApp_p$;

$p\phi SignOrD$ – вероятность отчуждения компонентов криптозащиты (ЭП), контролирующих целостность оригинала Элд (OrD).

4. Математическая модель отчуждаемости нормализованной копии электронного документа

Кроме оригинала документа (3) необходимо оценить отчуждаемость также и других семантических блоков модели (1). В первую очередь после оригинала Элд критична отчуждаемость нормализованной копии. Это обусловлено тем, что вероятность через десятилетия интерпретировать нормализованную копию выше, чем у документа в оригинальном формате. В первую очередь из-за того, что форматы долговременного хранения теоретически более интерпретируемы. В работе [4] подробно расписана модель нормализованной копии Элд. Согласно этой модели оценку отчуждаемости можно представить в следующем виде:

$$p\phi OdfD = (\sum_{(i=1,N8)} kN8_i (p\phi OdfDoc_i^{kN81} (\sum_{(l=1,L1)} kL1_{il} p\phi OdfLinkPic_{il})^{kN82}))^{k8} (\sum_{(j=1,N9)} kN9_j p\phi OdfPic_j)^{k9} (\sum_{(k=1,N10)} kN10_k p\phi Sign_k)^{k10}, \quad (4)$$

где весовые коэффициенты важности, назначаемые экспертами или автоматически, в простейшем случае равные между собой, должны удовлетворять условиям:

$$\sum k_i = 1, i=[8,10],$$

$$\sum_{(i=1,N8)} kN8_i = 1, \sum_{(l=1,L1)} kL1_{il} = 1, kN81 + kN82 = 1,$$

$$\sum_{(i=1,N9)} kN9_i = 1, \sum_{(i=1,N10)} kN10_i = 1;$$

$p\phi OdfDoc$ – вероятность отчуждения (нормализованного) текстового содержимого семантических блоков [1–N8] оригинала Элд; подробнее о форматах долговременного хранения см. [8]; семантические блоки $OdfDoc$ могут содержать множества ссылок на все графические материалы $OdfLinkPic$, вероятность их отчуждения – $p\phi OdfLinkPic$ для ссылок [1–L1];

$p\phi OdfPic$ – вероятность отчуждения множества [1–N9] нормализованной графической информации (растровые и векторные изображения, элементы презентаций и др.), подлежащей преобразованию из исходных Элд в графические форматы долговременного хранения; подробнее о форматах см. [8, 9];

$p\phi Sign$ – вероятность отчуждения множества ЭП [1–N10], заверяющих нормализованный документ (содержит в себе сертификаты подписавших, цепочку сертификатов, сертификаты удостоверяющих центров (УЦ), возможно СОС), см. подробнее [4].

Важно, что для нормализованного документа нужно сохранять как текстовые материалы в формате долговременного хранения, так и внешний вид в форматах долговременного хранения изображений.

5. Разработка математической модели отчуждаемости метаданных

Разработка математической модели для вероятности отчуждения метаданных $p\phi DMD$ требует отдельного исследования, которое автор планирует провести в ближайшем будущем.

В общем случае математическая модель будет сильно зависеть от состава метаданных Элд. В простейшем случае – это будет математическая модель отчуждаемости 13 реквизитов в основном или 18 в расширенном наборе метаданных стандарта [6].

Для полной модели необходимо добавить модели содержания документа, визуальной формы представления документа, различных выписок из журналов инвентаризации (аутентичности, интерпретируемости), журналов операций, журналов безопасности и др., а также математические модели отчуждаемости индексов, в том числе полнотекстовых.

6. Разработка математической модели отчуждаемости связанных данных

Нормативно-справочная информация (НСИ), т.е. классификаторы, словари, нормативные документы (CLI), другие Элд (LDI), на которые ссылается Элд, также важны для интерпретации Элд

в будущем. Если Элд, подлежащий долговременному хранению, ссылается на определенную НСИ или другие Элд, то эти данные должны также сохраняться вместе с Элд. Причем НСИ и связанные Элд должны сохраняться именно в той версии, которая была актуальна на момент создания Элд. НСИ и связанные Элд могут изменяться со временем, что приведет к проблеме неправильной интерпретации Элд.

Тогда математическую модель отчуждаемости НСИ можно представить в следующем виде:

$$p\phi CLI = \sum_{(i=1, N11)} kN11_i p\phi CLI_i \quad (5)$$

где $p\phi CLI_i$ – вероятность отчуждения связанной НСИ;

$kN11_i$ – коэффициент важности отдельного элемента (семантического блока) НСИ, причем $\sum_{(i=1, N11)} kN11_i = 1$. При отсутствии явного предпочтения между элементами НСИ, $kN11_i = (1/N11)$.

Основная проблема использования НСИ – автоматизация классификация Элд. Для ее решения существует несколько подходов: первый заключается в написании правил отнесения документов к классам, второй – в использовании машинного обучения.

На практике наиболее разумным оказывается комбинирование обоих подходов к решению проблемы автоматизации отнесения электронных документов к классам. Подробнее о принципах построения и обучения классификаторов см. [10].

Аналогично отчуждаемости НСИ, можно предложить следующую модель отчуждаемости связанных Элд:

$$p\phi LDI = \sum_{(i=1, N12)} kN12_i p\phi LDI_i, \quad (6)$$

где $p\phi LDI_i$ – вероятность отчуждения связанного Элд;

$kN12_i$ – коэффициент важности отдельного связанного Элд, причем $\sum_{(i=1, N12)} kN12_i = 1$. При отсутствии явного предпочтения между связанными Элд, $kN12_i = (1/N12)$.

7. Пример шкалы значений для оценки вероятности отчуждения

В качестве математической модели отчуждения Элд использована вероятностная модель. Но, как было сказано выше, вероятность здесь понимается как риск неотчуждения. Как можно его оценить.

Если испытания Бернулли в классическом понимании оценки вероятности невозможны или трудоемки, то можно использовать вербально-числовые значения оценки риска для приведенных

Табл. 1

Пример вербальной шкалы для приблизительной оценки вероятности отчуждения

Номер п/п	Наименование значений	Числовой эквивалент
1	Не извлекается	0
2	Скорее не извлекается	0,25
3	Частично извлекается	0,5
4	Скорее извлекается	0,75
5	Полностью извлекается	1

выше показателей математической модели. Подробнее см. [7].

Оценка по данной шкале выполняется экспертом (группой экспертов), т.е. лицом (лицами) принимающим решение (ЛПР), непосредственно осуществляющими организацию долговременного хранения ЭД. ЛПР должны быть знакомы с программно-аппаратной средой долговременного хранения. Пример шкалы значений приведен в табл. 1.

Для более точной оценки вероятности отчуждения ЭД в программное обеспечение долговременного хранения должна быть заложена функция периодической автоматической инвентаризации фонда ЭД с целью автоматического определения возможности отчуждения каждого ЭД от программно-аппаратной среды хранения с автоматическим подсчетом вероятности извлечения (имитация испытаний Бернулли).

Если отчуждение ЭД (или какого-либо семантического блока ЭД) автоматически выполнить невозможно персонал, обслуживающий программно-аппаратную среду хранения, должен быть уведомлен о нарушении отчуждения конкретного ЭД. Также должна быть предоставлена возможность проверки проблемного ЭД на возможность его отчуждения в ручном режиме. Если и эта попытка не удалась, показатель вероятности извлечения данного ЭД (или какого-либо семантического блока ЭД) автоматически вычисляется системой ЭА как равный нулю.

Заключение

В данной статье предложена математическая модель отчуждаемости ЭД от программно-аппаратной среды хранения. Она тесно связана с предложенной автором ранее математической моделью ЭД долговременного хранения (см. [4]). Предложенная математическая модель вероятностная. Однако, т.к. испытания Бернулли в классическом понимании оценки вероятности невозможны или трудоемки, то в статье показано, как можно использовать оценки риска для приведенных показателей математической модели.

Математическая модель построена с учетом того, что среда хранения ЭД всегда подвижна и изменчива, подвержена параметрическим возмущениям. Следовательно, при долговременном хранении ЭД должен быть, как объект управления, максимально стабилизирован по основной характеристике – сохранность. Для этого он, в частности, должен быть максимально независим от среды хранения. В качестве меры независимости предложена математическая модель оценки риска неотчуждаемости документа от среды хранения.

Предложены математические модели оценки отчуждаемости семантических блоков оригинала ЭД, нормализованной копии ЭД, связанных НСИ и других ЭД. В дальнейшем планируется разработка математической модели оценки отчуждаемости метаданных ЭД.

Выполненные расчеты и практическое применение предложенных в статье подходов в ряде проектов электронного архива – электронные архивы для Пенсионного Фонда РФ, ООО «Юридическая фирма Городисский и Партнеры», ООО «Когнитивные технологии» – дают основания полагать адекватность представленной модели условиям задачи долговременного хранения.

Литература

1. *Solovyev A. V.* The Problem of Defining the Concept of “Electronic Document for Long-Term Storage” // *Lecture Notes in Networks and Systems*. 2023. vol.597(2), pp.326-333. doi: 10.1007/978-3-031-21438-7_26.
2. *Емельянов Н.Е.* Виды представления структурированных данных // *Теоретические основы информационной технологии. Сборник трудов ВНИИСИ*. № 22, М.:ВНИИСИ. 1988. С.42–46.
3. *Емельянов Н.Е.* Теоретический анализ документного интерфейса: Препринт / Н.Е. Емельянов. М.: Всесоюзный научно-исследовательский институт системных исследований. 1987. 40 с.
4. *Соловьев А.В.* Математическая модель электронного документа долговременного хранения

- ния // Информационные технологии и вычислительные системы. №2. 2022. С.30-36. doi: 10.14357/20718632220204.
5. Solovyev A.V. Long-Term Digital Documents Storage Technology // Lecture Notes in Electrical Engineering. 2020. vol.641. P. 901-911. doi: 10.1007/978-3-030-39225-3_97.
 6. ГОСТ Р 7.0.10-2019 (ИСО 15836-1:2017) Система стандартов по информации, библиотечному и издательскому делу. НАБОР ЭЛЕМЕНТОВ МЕТАДААННЫХ “ДУБЛИНСКОЕ ЯДРО”. Основные (ядерные) элементы. (System of standards for information, librarianship and publishing. The Dublin Core metadata element set. Basic (core) elements).
 7. Теория принятия решений // Петровский А.Б. – М.: Академия. 2009. 400 с.
 8. Николаев Д.П., Постников В.В., Усилин С.А. Cognitive PDF/A – технология оцифровки текстовых документов для публикации в Интернет и долговременного архивного хранения // Труды ИСА РАН. 2009. Т. 45. С. 159–173.
 9. Берестова В.И. Средства и методы создания электронного документа, содержащего графические образы // Журнал «Делопроизводство». 2014. № 1. С.45–56.
 10. 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.

Соловьев Александр Владимирович, Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия. Главный научный сотрудник. Доктор технических наук. Область научных интересов: системный анализ, системы управления базами данных, теория надежности, математическое моделирование, долговременное хранение электронных документов. E-mail: soloviev@isa.ru

Mathematical model of the alienability of an electronic document from the storage environment

A.V. Solovyev

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, Moscow, Russia

Abstract. The article proposes a mathematical model of the alienability of an electronic document from a software and hardware storage environment. The storage medium of an electronic document is always mobile and changeable, subject to parametric disturbances. Therefore, during long-term storage, an electronic document, as a control object, should be maximally stabilized in terms of the main characteristic - safety. To do this, in particular, it should be as independent as possible from the storage environment. As a measure of independence, a mathematical model for assessing the risk of inalienability of a document from the storage environment is proposed. The performed calculations and practical application in a number of electronic archive projects give reason to believe that the presented model is adequate for the conditions of the problem of long-term storage.

Keywords: *long-term storage, document safety, electronic document, alienability, risk assessment.*

DOI: 10.14357/20790279230401 **EDN:** LH CYGP

References

1. Solovyev A.V. 2023. The Problem of Defining the Concept of “Electronic Document for Long-Term Storage”. Lecture Notes in Networks and Systems. 597(2): 326-333. doi: 10.1007/978-3-031-21438-7_26.
2. Emelyanov N.E. 1988. Vidy predstavleniya strukturirovannykh 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.
3. 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.
4. Solovyev A.V. 2022. Matematicheskaya model elektronnoho dokumenta dolgovremenno-

- go khraneniya [Mathematical model of an electronic document of long-term storage] // *Informatsionnyye tekhnologii i vychislitel'nyye sistemy* [Information Technology and Computing Systems] 2: 30-36. doi: 10.14357/20718632220204.
5. *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.
 6. GOST R 7.0.10-2019 (ISO 15836-1:2017) System of standards for information, librarianship and publishing. The Dublin Core metadata element set. Basic (core) elements.
 7. *Petrovsky A.B.* 2009. *Theoriya prinyatiya resheniy* [Decision theory]. M.: Akademy. 400 p.
 8. *Nikolayev D.P., Postnikov V.V., and Usilin S.A.* 2009. Cognitive PDF/A – tekhnologiya otsifrovki tekstovykh dokumentov dlya publikatsii v Internet i dolgovremennogo arkhivnogo khraneniya [Cognitive PDF / A - technology for digitizing text documents for publication on the Internet and long-term archival storage]. *Trudy ISA RAN* [Proceedings of the ISA RAS]. 45: 159–173.
 9. *Berestova V.I.* 2014. Sredstva i metody sozdaniya elektronnoho dokumenta, soderzhashchego graficheskiye obrazy [Means and methods for creating an electronic document containing graphic images]. *Deloproizvodstvo* [J Office work]. 1: 45–56.
 10. *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.

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