

Задача генерации выходных форм документов информационных систем

А.В. СОЛОВЬЕВ

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

Аннотация. В статье формализована задача генерации выходных (отчетных) форм документов по динамически изменяющимся структурам хранения данных в информационных системах. Предложен подход к автоматизации процесса генерации выходных форм. Рассмотрены структуры данных, необходимые и достаточные для представления цифровизации информационных систем в выходных (отчетных) документах. Преимуществом предложенного подхода является упрощение и существенное ускорение процесса генерации выходных форм, который в настоящее время очень трудоемок.

Ключевые слова: генерация выходных форм документов, структуры хранения данных, структуры данных выходных (отчетных) документов.

DOI: 10.14357/20790279240303 **EDN:** GMDPNM

Введение

В современном мире, охваченном процессами цифровой трансформации, когда, например, цифровые формы становятся ключевым фактором производства (см. [1]), а информационные системы, использующие и обрабатывающие эти данные, становятся неотъемлемой частью производства, крайне важно удержать инициативу и технологическое первенство. В частности, что касается разработки информационных систем (ИС), крайне важной задачей является их быстрая разработка и модификация. Дело в том, что в момент существенных и быстрых цифровых преобразований экономики, перехода на новые технологии, вопрос времени становится крайне важным. В такой ситуации невозможно разрабатывать и модернизировать ИС годами. Скорость разработки и модификации будет решающей в конкурентной борьбе за технологическое преимущество и, как следствие, за устойчивость и конкурентоспособность экономики в целом (см, например, [2-4]).

Конечно, вопросы ускорения разработки и модификации ИС крайне обширны. В данной статье автор рассматривает только два аспекта, а именно:

- быстрое конструирование ИС и структур хранения данных на основе форм входных документов;
- генерация выходных (отчетных) форм документов по динамически изменяющимся структурам хранения данных.

Основной упор в статье сделан на втором аспекте, т.к. первый аспект в настоящее время так или иначе решен, чему посвящен далее краткий обзор.

Ускорение же второго аспекта до настоящего времени остается задачей нерешенной. Формализации данной задачи, рассмотрению структур данных выходных документов, подходу к решению задачи автоматизации генерации выходных форм документов, посвящена данная статья.

1. Краткий обзор источников и проблем быстрой разработки информационных систем

Из краткого обзора литературы следует, что проблема быстрого создания ИС крайне актуальна, решения по созданию конструкторов ИС создавались и создаются. При этом используется огромное количество разных подходов, вплоть до использования методов искусственного интеллекта (ИИ) для создания хранилищ данных для неструктурированных и слабоструктурированных документов и массивов информации.

Только обзор работ за последние годы позволяет создать следующую картину решений и направлений исследования.

Важным направлением при цифровизации экономики является автоматизация процесса соз-

дания гибких производственных системы (см., например, [5]). В частности, в работах описан процесс автоматизации создания структур хранения 2D и 3D данных (баз данных, БД) для таких систем. Используемый подход основан на том, что процесс создания ИС строится от структур входных данных, на основании которых создаются структуры реляционной БД под конкретные производственные задачи.

В работах [6, 7] представлен альтернативный подход, согласно которому создается платформенное решение, позволяющее настраиваться на разные структуры БД и извлекать из них необходимую информацию, в том числе на основе онтологий (см. [7]). Ограничением данного подхода является то, что он может использоваться в достаточно узких предметных областях, где известны заранее параметры, которые требуется извлечь из БД для выполнения функций ИС. Более того, параметры БД, как правило, стандартизованы нормативными документами. Однако такой подход, безусловно, имеет право на существование и практическую ценность.

Отдельных исследований требуют подходы, связанные с автоматизацией процесса построения структур хранения данных в ИС на основе слабоструктурированных и неструктурированных документов. Магистральным направлением в настоящее время здесь является использование методов ИИ для создания БД таких ИС (см., например, [8, 9]). Подход достаточно перспективный, но требующий создания достаточно сложных структур метаданных для своей реализации и нечеткого поиска реквизитов в неструктурированных текстах.

Нельзя не отметить также наличие подходов, связанных с использованием RAD-технологий (RAD – Rapid Application Development) и CASE-средств (см., например, [10-12]). Хотя сами концепции использования RAD и/или CASE возникли еще в конце 80-х годов XX века. Эта концепция до сих пор остается достаточно востребованной. Однако ограничения CASE-средств, в том числе отсутствие универсальности и дороговизна, ресурсоемкость RAD-технологий создает определенные сложности с их массовым внедрением. К тому же идея автоматической генерации программного кода (программирование без программирования) в настоящее время также имеет свои ограничения. Одно дело создать автоматически программы по заданному алгоритму. С этим сейчас справляются средства генеративного интеллекта типа ChatGPT, другое дело автоматическая генерация программ по заданным параметрам бизнес-процессов, требую-

ющих универсальности описания. Хотя здесь тоже есть заметные успехи, несмотря на то, что предложенный в [13] подход натывается на ограничения самой методологии IDEF0, например, отсутствие ветвлений и условных переходов.

И, наконец, еще одним подходом к решению проблемы быстрого создания ИС и структур их БД является разработка достаточно универсальных платформенных систем (см., например, [14, 15]). Достоинством такого является достаточно универсальный продукт, который можно применять в различных предметных областях. Недостатками являются, как правило, большой объем ИС, наличие избыточной функциональности в рамках конкретных применений, недостаточная гибкость для настройки под вкусы и требования конкретных пользователей.

Необходимо отметить, что в Отделе 94 ФИЦ ИУ РАН также был разработан прототип решения для быстрого создания ИС на основе подхода быстрого построения и модификации структур БД ИС на основе входных форм документов и создания платформенного универсального решения. Подробному описанию разработанного прототипа решения будет посвящена отдельная статья.

В качестве вывода для данного раздела необходимо отметить, что в рамках перечисленных подходов в целом решена проблема автоматизации создания и модификации структур БД на основе входных документов и разработки достаточно универсального платформенного решения. Однако до сих пор нерешенной остается задача автоматизации изменения форм выходных документов, созданных в ИС на основе структур БД. В случае изменения структур БД требуется изменения форм выходных документов (отчетов) вручную.

Эта задача не решена, даже несмотря на очень мощное развитие средств генерации отчетов и BI-систем. Эти системы позволяют создавать сложные отчеты с развитой графикой, обобщением данных, снабжены инструментами анализа данных (диаграммы, сводные отчеты, кросс-табличные отчеты и проч.). Однако изменение структур данных требует их перестроения оператором вручную. По мнению автора, для успешного решения задачи необходимо ее правильно формализовать, а также проанализировать структуры данных, необходимые для представления в выходных документах. Этим вопросам и посвящена данная статья.

2. Формализация задачи исследования

В общем виде задачу исследования можно сформулировать так. На вход информационной

системы поступает множество входных данных. В общем случае входные данные имеют структуру явно или неявно выраженную в их форме. ИС при вводе данных представляет их в виде некоторой структуры, чаще выраженной шаблонами входных форм данных, предназначенных для ввода. По формам входных данных ИС осуществляет создание структур для их хранения в своей БД. При изменении входных данных, изменяются их формы, после чего происходит автоматическое или автоматизированное перестроение структур хранения данных. Задача перестроения структур хранящихся данных, согласно выше проведенному обзору, в целом решена. Затем по хранящимся данным, соответственно их структурам хранения, создаются формы выходных документов ИС (отчетные формы). Необходимо разработать механизмы, методы и средства автоматического (автоматизированного) преобразования форм выходных документов ИС при изменении структур хранения данных. Задача такого преобразования пока не решена, что приводит к тому, что на преобразование форм (шаблонов) выходных документов тратится много времени при сопровождении и эксплуатации ИС.

В формализованном виде задачу исследования можно представить так.

Дано:

- 1) Множество входных данных ИС $D = \{d_i\}$.
- 2) Множество форм входных документов ИС для ввода данных $FID = \{fid_i\}$.
- 3) Множество структур хранения данных ИС $DSS = \{dss_i\}$.
- 4) Множество, хранящихся согласно DSS , данных ИС $ISD = \{isd_i\}$.
- 5) Множество способов автоматического преобразования FID в DSS : $IAT: FID \rightarrow DSS$.
- 6) Множество способов автоматического преобразования D в ISD : $IDAT: D \rightarrow ISD$.
- 7) Множество форм выходных документов ИС $FOD = \{fod_i\}$.
- 8) Множество создаваемых по FOD выходных документов $OD = \{od_i\}$.

Найти:

- 1) Множество способов автоматического преобразования DSS в FOD : $OAT: DSS \rightarrow FOD$.
- 2) Множество способов автоматического преобразования ISD в OD : $ODAT: ISD \rightarrow OD$.

Для того, чтобы решить поставленную задачу исследования, необходимо для начала проанализировать структуры данных DSS и FOD .

Структуры хранения данных, как правило, представлены в виде плоских двумерных реляционных таблиц, связанных отношениями 1:N, 1:1,

N:N. Как правило, тип отношений 1:N существенно превалирует над остальными.

Анализ же структур данных FOD требует более детального изучения.

3. Анализ структур данных выходных документов и их форм

Выходной документ (od_i) – это, как правило, отчет, представленный на экране или напечатанный на бумаге. Все современные генераторы отчетов ориентированы на представление данных на плоском бумажном носителе. Для этого в них реализована возможность размещать структуры данных на листах бумаги, как в обычных текстовых редакторах офисных приложений.

В таком случае будет справедливым утверждение о представлении такого выходного документа (od_i) и его формы (шаблона) (fod_i) в виде совокупности семантических блоков, как это сделано в работе автора [16]. Т.е. $OD = \{od_i\}$, $od_i = \bigcup_{(j=1..N)} (B_j)$, каждому od_i из множества OD поставим в соответствие некоторый граф $G(od_i) = (V, E)$, представляющий схему его семантических блоков B_j . V – множество вершин графа, E – множество дуг. $V = \{B_j\}$. $E = \{(B_j, B_k): B_j \rightarrow B_k\}$. Аналогично представим и fod_i .

Будем считать, что od_i (fod_i) разбивается на семантические блоки, если:

- $B_j \cap B_k \neq \emptyset$, то $B_j \subseteq B_k$ или $B_k \subseteq B_j$;
- $\exists B_k \subseteq B_j$, то $B_j = \bigcup B_k$ для всех $B_k \subseteq B_j$.

Семантический блок B_j называется простым (или терминальным), если он не содержит никаких внутренних блоков. На графе семантических блоков $G(od_i)/G(fod_i)$ – это листовые (не имеющие потомков) вершины V .

Назовем рангом $r(B_j)$ семантического блока B_j длину максимальной цепочки вложенных блоков с корнем в B_j . Т.е. $r(B_j) = \max \{m: \exists \{B_{j_1}, \dots, B_{j_m}\}, \text{ что } B_{j_m} \subset \dots \subset B_{j_1} \subset B_j\}$. Если B_j простой, то $r(B_j) = 0$. Иными словами, на графе $G(od_i)$ – это длина ветки с корнем в B_j «минус» единица. С помощью признака $r(B_j) = 0$ удобно пометать листовые вершины (простые блоки).

Назовем повторяемостью $s(B_j)$ семантического блока B_j количество повторений (возможно с разными данными) в od_i . Если блок не повторяется, $s(B_j) = 0$. Т.к. при преобразовании ($ODAT: ISD \rightarrow OD$) выполняется запрос к БД, то, по сути, количество повторов – это количество возвращаемых строк запроса (результат запроса к реляционной БД – всегда плоская двумерная таблица) «минус» 1.

Выходные (отчетные) документы OD и формы FOD , на основании которых эти документы

создаются, представляют данные, хранящиеся в ИС согласно **DSS**, в виде гладкого текста, таблиц, нумерованных или маркированных списков, изображений, рисунков, графических фигур. Частным случаем текстового представления также являются сноски и ссылки.

При этом гладкий текст – это последовательность входящих в него данных B_j . Таблицы – табличное представление данных B_j . Нумерованные и маркированные списки – иерархическое (в общем случае) представление данных B_j . Изображения, рисунки, графические элементы и прочее – контурное представление B_j , характеризующиеся областью изображения с контуром произвольной формы. Все остальные представления данных B_j можно характеризовать как смешанное представление.

Тогда можно утверждать, что все данные B_j представимы в структурах **OD** и **FOD** в виде:

- последовательностей;
- таблиц;
- иерархий;
- контуров;
- смешанное (из указанных выше четырех видов представлений).

Можно утверждать также, что:

- B_j представим в виде последовательности, если $r(B_k) = 0$ для всех $B_k \subseteq B_j$;
- B_j представим в виде таблицы, если $s(B_k) > 0$ для всех $B_k \subseteq B_j$;
- B_j представим в виде иерархии, если $r(B_j) > 0$.

На основании приведенных рассуждений можно утверждать о применимости методов теории множеств, теории графов для организации структур данных **OD** и **FOD**, нацеленных на независимость от реализации БД средствами конкретной СУБД, использующейся в ИС.

Теперь перейдем к описанию подхода к автоматизации процесса генерации выходных форм документов **FOD** и выходных документов **OD** по этим формам.

4. Подход к автоматизации процесса генерации выходных форм и документов

Автоматизацию процесса генерации выходных форм (**FOD**) и документов на их основе (**OD**) логично разбить на два этапа.

Первый этап – собственно автоматизация генерации выходных форм (**FOD**) на основе динамически меняющихся структур хранения данных (**DSS**).

Второй – генерация выходных документов (**OD**) на основе изменившихся форм (**FOD**).

Второй этап гораздо проще поддается автоматизации, т.к. в настоящее время существует достаточно большое количество программных средств, позволяющих создавать выходные документы (отчеты) на основе шаблонов. Такие средства относятся либо к классу генераторов отчетов, либо к классу BI-средств (BI, Business Intelligence), т.е. средств аналитической обработки информации.

Как генераторы отчетов, так и BI-средства обладают развитой функциональностью, позволяющей автоматизировать не только представление данных, хранящихся согласно структурам хранения (**DSS**), но и производить их обработку: суммирование и другие подсчеты, группировки, сортировки, графическое представление в виде графиков, диаграмм, кросс-таблиц и прочее. К средствам генераторов отчетов и BI-средствам можно отнести: Oracle BI Publisher, Crystal Reports, SQL Server Reporting Services, Microsoft Power BI, Qlik, Icube, CyberQuery, 1C:Аналитика и др. К сожалению, формат статьи не позволяет сделать полный обзор данных средств, тем более, что подобные обзоры периодически появляются (см., например, [17]).

В целом же отметим, что в зависимости от нужной функциональности можно использовать то или иное средство при создании ИС. В целом же можно отметить, что указанные выше средства в целом решают проблему поиска способов автоматического преобразования **ISD** в **OD**: **ODAT: ISD → OD**.

Совсем по-другому обстоит дело с первым этапом, т.е. автоматизацией генерации выходных форм (**FOD**) на основе динамически меняющихся структур хранения данных (**DSS**). В настоящее время не существует таких средств, поэтому нужен подход к поиску **OAT: DSS → FOD**.

Если чуть шире описать проблему динамически изменяющихся данных в информационной системе, опирающейся на входные данные (**D**) и формы входных данных (**FID**), то общий алгоритм функционирования ИС будет выглядеть примерно так:

- изменение структуры входных данных (**D**);
- корректировка форм входных документов ИС для ввода данных (**FID**);
- проверка соответствия **FID** и **DSS**;
- автоматическое преобразование **FID** в **DSS**: **IAT: FID → DSS**;
- сохранение истории изменений **DSS**;
- автоматическое преобразование **D** в **ISD**: **IDAT: D → ISD**;
- проверка соответствия **FOD** и **DSS**;
- автоматическое преобразование **DSS** в **FOD**: **OAT: DSS → FOD**;

– автоматическое преобразование *ISD* в *OD*: *ODAT: ISD* → *OD*.

Поскольку выше определена структура данных выходных документов и их форм (*OD* и *FOD*), как граф их семантических блоков B_j , то логично представлять модель соответствия *FOD* и *DSS* как графовую математическую модель. В каждом узле (B_j) этой модели хранится информация о соответствии структур *FOD* и *DSS*, а также метаданные о представлении узлов (B_j) в выходном документе (*OD*), т.е. полиграфическая спецификация B_j . Назовем такую модель – моделью представления *FOD*.

Тогда ИС при изменении структур хранения данных *DSS* (выполнения преобразования *IAT: FID* → *DSS*) выполняет проверку соответствия структур *FOD* и *DSS* с помощью промежуточной модели представления *FOD*. При несовпадении структур осуществляется просмотр истории изменений *DSS*. Потом на основании ретроспективы изменений следует анализ возможных замен имен, изменений размеров и топов полей структуры *DSS*. Возможно, для анализа истории изменений понадобятся методы, использующиеся в системах искусственного интеллекта, такие как нечеткий поиск [18], т.к. в его основе графовые модели представления данных для поиска аналогов в структурах *DSS*.

Поскольку моделью представления *FOD* – это графовая модель, узлами которой могут быть множества, можно утверждать о применимости методов теории множеств и графов, комбинаторного поиска как для организации структур данных *FOD*, так и для организации поиска изменений и проверки соответствий между *FOD* и *DSS*. Кроме того, преобразование графов промежуточной модели также, несомненно, будет лежать в основе способов автоматического преобразования *DSS* в *FOD*: *OAT: DSS* → *FOD*.

Заключение

В данной статье выполнен краткий обзор источников и проблем быстрой разработки информационных систем. Формализована задача генерации выходных (отчетных) форм документов по динамически изменяющимся структурам хранения данных в информационных системах.

Предложены и рассмотрены структуры данных, необходимые и достаточные для представления цифровых данных информационных систем в выходных (отчетных) документах. Сделан вывод о применимости методов теории множеств и графов, комбинаторного поиска как для организации структур данных и выходных документов, так и для их хранения.

Предложен подход к автоматизации процесса генерации выходных форм. Показана необходимость использования промежуточной графовой модели для анализа соответствия структур данных хранения и форм выходных документов и автоматизации генерации выходных форм. Преимуществом предложенного подхода является упрощение и существенное ускорение процесса генерации выходных форм, который в настоящее время достаточно трудоемок из-за необходимости проведения проверки и выполнения модернизации форм выходных документов вручную.

Литература

1. Программа «Цифровая экономика Российской Федерации». Утверждена распоряжением Правительства Российской Федерации от 28 июля 2017 г. № 1632-р. М.: 2017. 88 с.
2. Bartodziej C.J. The Concept Industry 4.0 // Springer Fachmedien Wiesbaden: Wiesbaden, Germany. 2017. P. 27–50.
3. Schilz A., Rehbein M. Digitization of Cultural Heritage // Handbook Industry 4.0: Law, Technology, Society; Springer: Berlin, Heidelberg. 2022. P. 1193–1212.
4. Aheleroff S., Huang H., Xu X., Zhong R.Y. Toward sustainability and resilience with Industry 4.0 and Industry 5.0 // Frontiers in Manufacturing Technology. 2022. 2. 951643. <https://doi.org/10.3389/fmtec.2022.951643>.
5. Abdullayev G. Algorithmically and programming management of design process of flexible manufacture system // Системная инженерия и информационные технологии. 2021. Т. 3, № 1 (5). С. 20–28.
6. Арзучинцев А.В., Кедрин В.С., Кедрина М.С. Архитектура иерархически модифицируемо-пересекающейся базы данных экологических параметров // Вестник Бурятского государственного университета. Математика, информатика. 2022. №1. С. 3-17.
7. Taberko V., Ivaniuk D., Zotov N., Orlov M., Pupena O., Lutska N. Principles of building a system for automating the activities of a process engineer based on an ontological approach within the framework of the Industry 4.0 concept // Открытые семантические технологии проектирования интеллектуальных систем. 2021. № 5. С. 209-218.
8. Вотинова Е.С., Чуприна С.И. Автоматизация построения баз данных из неструктурированных текстовых документов // Математика и междисциплинарные исследования – 2021. [Электронный ресурс]: материалы Всероссийской научно практической конференции молодых

- ученых с международным участием (г. Пермь. 18–20 октября 2021 г.) / гл. ред. А. П. Шкарапу-та; Пермский государственный национальный исследовательский университет. Пермь, 2021. – С. 119-122. URL: www.psu.ru/files/docs/science/books/sborniki/mmi-2021.pdf. (дата обращения 15.03.2024).
9. Пальмов С.В., Шамрин Д.А. Функциональная структура «Фермы данных» // Евразийское Научное Объединение. 2021. № 5-1 (75). С. 42-45. DOI: 10.5281/zenodo.4926581.
 10. Кондратьев В.Ю., Кутыкова Е.М. Прототипное проектирование ЭИС (RAD-технология) // Информационное общество: современное состояние и перспективы развития. Сборник материалов XIV международного форума. Краснодар. 2021. С. 429-431.
 11. Lyashenko I.I. About the use of CASE-technologies in the process of designing information systems // Вестник Инновационного Евразийского университета. 2022. № 2 (86). С. 126-133.
 12. Лайпанов А.А., Гебенев Р.М. Проектирование информационных систем с использованием CASE-технологий // Тенденции развития науки и образования. 2021. № 80-2. С. 118-121. DOI: 10.18411/trnio-12-2021-89.
 13. Басманов С.Н., Вакалюк А.А. Разработка подхода к созданию гибкого пользовательского интерфейса на основе преобразования IDEF0-диаграммы // Современные наукоемкие технологии. 2020. № 5. С. 20-25.
 14. Черныш Б.А., Картамышев А.С. Разработка ядра интегрированной информационной системы // Программные продукты и системы (Software & Systems). 2021. № 2. С. 237-244. DOI: 10.15827/0236-235X.134.237-244.
 15. Автоматизированная система управления предприятием // Рули24. [Электронный ресурс]: URL: <https://ruli24.ru/> (дата обращения 15.03.2024).
 16. Соловьев А.В. Математическая модель электронного документа долговременного хранения // Информационные технологии и вычислительные системы. №2. 2022. С.30-36. DOI: 10.14357/20718632220204.
 17. King Tim. The 19 Best Data Catalog Tools and Software for 2024 // Solutions Review. 2024. URL: <https://solutionsreview.com/data-management/the-best-data-catalog-tools-and-software/> (дата обращения 19.03.2024).
 18. Kang Bo-Yeong, Dae-Won Kim, Hae-Jung Kim. Fuzzy Information Retrieval Indexed by Concept Identification // Lecture Notes in Computer Science, vol. 3658, Springer Berlin / Heidelberg. 2005. P. 179–186. DOI:10.1007/11551874_23.

Соловьев Александр Владимирович. 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

The problem of generating output forms of information systems documents

A.V. Solovyev

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

Abstract. The article formalizes the task of generating output (reporting) forms of documents based on dynamically changing data storage structures in information systems. An approach to automating the process of generating output forms is proposed. The data structures necessary and sufficient for representing digital data of information systems in output (reporting) documents are considered. The advantage of the proposed approach is the simplification and significant acceleration of the process of generating output forms, which is currently quite labor-intensive.

Keywords: *generation of output document forms, digital data, data structures of output (reporting) documents.*
DOI: 10.14357/20790279240303 **EDN:** GMDPNM

References

1. The Government of the Russian Federation. 2017. PROGRAM "Digital Economy of the Russian Federation" approved by the order of at July 28, 2017. № 1632-р: 88.
2. *Bartodziej C.J.* 2017. The Concept Industry 4.0. Springer Fachmedien Wiesbaden: Wiesbaden, Germany, pp. 27–50.
3. *Schilz A., Rehbein M.* 2022. Digitization of Cultural Heritage. Handbook Industry 4.0: Law, Technology, Society; Springer: Berlin, Heidelberg, pp. 1193–1212.
4. *Aheleroff S., Huang H., Xu X., Zhong R.Y.* 2022. Toward sustainability and resilience with Industry 4.0 and Industry 5.0. *Frontiers in Manufacturing Technology*, 2, 951643. <https://doi.org/10.3389/fmtec.2022.951643>.
5. *Abdullayev G.* 2021. Algorithmically and programming management of design process of flexible manufacture system. *Systems Engineering and Information Technology*, vol. 3, № 1 (5): 20–28.
6. *Arguchintsev A.V., Kedrin V.S., Kedrina M.S.* 2022. Architecture of a hierarchically modified-intersecting database of environmental parameters. *Bulletin of Buryat State University. Mathematics, computer science*, №1: 3-17.
7. *Taberko V., Ivaniuk D., Zotov N., Orlov M., Pupena O., Lutska N.* 2021. Principles of building a system for automating the activities of a process engineer based on an ontological approach within the framework of the Industry 4.0 concept // Open semantic technologies for designing intelligent systems. № 5: 209-218.
8. *Votinova E.S., Chuprina S.I.* 2021. Automation of database construction from unstructured text documents. *Mathematics and interdisciplinary research - 2021: materials of the All-Russian scientific and practical conference of young scientists with international participation* (Perm, October 18–20, 2021) / Ch. ed. A. P. Shkaraputa; Perm State National Research University. Perm: 119-122. URL: www.psu.ru/files/docs/science/books/sborniki/mmi-2021.pdf. (Accessed 15.03.2024).
9. *Palmov S.V., Shamrin D.A.* 2021. Functional structure of the Data Farm. *Eurasian Scientific Association*, № 5-1 (75): 42-45. DOI: 10.5281/zenodo.4926581.
10. *Kondratyev V.Yu., Kutykova E.M.* 2021. Prototype design of EIS (RAD technology). *Information society: current state and development prospects. Collection of materials from the XIV International Forum*. Krasnodar: 429-431.
11. *Lyashenko I.I.* 2022. About the use of CASE-technologies in the process of designing information systems. *Bulletin of the Innovative Eurasian University*, № 2 (86): 126-133.
12. *Laipanov A.A., Gebenov R.M.* 2021. Design of information systems using CASE technologies. *Trends in the development of science and education*, № 80-2: 118-121. DOI: 10.18411/trnio-12-2021-89.
13. *Basmanov S.N., Vakalyuk A.A.* 2020. Developing an approach to creating a flexible user interface based on IDEF0 diagram transformation. *Modern science-intensive technologies*, № 5: 20-25.
14. *Chernysh B.A., Kartamyshev A.S.* 2021. Development of the core of an integrated information system. *Software & Systems*, № 2: 237-244. DOI: 10.15827/0236-235X.134.237-244.
15. Automated enterprise management system. Ruli24. URL: <https://ruli24.ru/> (Accessed 15.03.2024).
16. *Solovyev A.V.* 2022. Mathematical model of an electronic document for long-term storage. *Information technology and computing systems*, №2: 30-36. DOI: 10.14357/20718632220204.
17. *King Tim.* 2024. The 19 Best Data Catalog Tools and Software for 2024. *Solutions Review*. URL: <https://solutionsreview.com/data-management/the-best-data-catalog-tools-and-software/> (Accessed 19.03.2024).
18. *Kang Bo-Yeong, Dae-Won Kim, Hae-Jung Kim.* 2005. Fuzzy Information Retrieval Indexed by Concept Identification. *Lecture Notes in Computer Science*, vol. 3658, Springer Berlin / Heidelberg, pp. 179–186, DOI:10.1007/11551874_23.

Solovyev A.V. Chief Researcher, Doctor of Technical Sciences. Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia. E-mail: soloviev@isa.ru