Анатолий А. Чуприно
ФГБУ «Всероссийский научно-исследовательский институт радиоэлектроники»,
ул. Колпакова, 2а, Мытищи, Московская обл., 141002, Россия
e–mail: toliy1962@list.ru, https://orcid.org/0000-0002-8785-6797

МЕТОДИЧЕСКИЙ ПОДХОД К ОБЪЕДИНЕНИЮ ИНФОРМАЦИОННЫХ РЕСУРСОВ В СИСТЕМЕ РАСПРЕДЕЛЕННОЙ ОБРАБОТКИ ИНФОРМАЦИИ

DOI: http://dx.doi.org/10.26583/bit.2023.1.08

Аннотация. В статье рассмотрены вопросы, связанные с формированием начального информационного массива, представленного содержанием объединенных баз данных, как носителя информации. Методический подход к объединению информационных ресурсов в системе распределенной обработки информации, в отличие от известных методов интеграции информационных ресурсов, основан на процедуре переноса данных по техническим характеристикам с несовпадающими значениями атрибутов электронных компонентов, описанных в серии новых национальных стандартов. Уточняющим действием при объединении информационных массивов является процедура переноса данных по техническим характеристикам (ТХ) с несовпадающими значениями атрибута ЭК. Данная процедура формирует полное множество технических характеристик результирующего ресурса, а алгоритм в методе является новым механизмом на множестве существующих и используемых методов объединения данных, описанных в национальных стандартах ГОСТ РФ по выбору ТХ для электронной компонентной базы отечественного производства (ЭКБ ОП).Таким образом, особенностью разработанного метода является алгоритм объединения отраслевых данных и создания единого представления совокупности всех данных по ЭКБ в рамках одной предметной области. В статье описан и проанализирован процесс объединения данных из фрагментов двух информационных систем. Разработаны UML диаграммы деятельности и логические условия для выбора альтернативных процедур в этом алгоритме. Приведены случаи необходимости участия специалиста в процессе объединения данных. Использование данного подхода и нормативной базы в виде серии национальных ГОСТ Р 59988.XX.XX.XX-2022 позволит стандартизовать информационные ресурсы электронной компонентной базы отечественного производства и конструкторскую документацию, описывающую опытные образцы радиоэлектронной аппаратуры на всех этапах жизненного цикла.

Ключевые слова: информационные ресурсы, база данных, объединение информации, объединение данных, метод объединения данных, объединенное информационное пространство, аналитические вычисления, гиперкуб информации, система поддержки принятия решений.
Для цитирования: ЧУПРИНОВ, Анатолий А. МЕТОДИЧЕСКИЙ ПОДХОД К ОБЪЕДИНЕНИЮ ИНФОРМАЦИОННЫХ РЕСУРСОВ В СИСТЕМЕ РАСПРЕДЕЛЕННОЙ ОБРАБОТКИ ИНФОРМАЦИИ. Безопасность информационных технологий, [S.l.], т. 30, № 1, с. 108–122, 2023. ISSN 2074-7136. URL: https://bit.spels.ru/index.php/bit/article/view/1483. DOI: http://dx.doi.org/10.26583/bit.2023.1.08.

Anatoliy A. Chuprinov
Federal State Budgetary Institution "All-Russian Research Institute of Radio Electronics",
Kolpakova Str. 2a, Mytischi, 141002, Russia
e–mail: toliy1962@list.ru, https://orcid.org/0000-0002-8785-6797

Methodological approach to the integration of information resources in a distributed information processing system

DOI: http://dx.doi.org/10.26583/bit.2023.1.08

Abstract. The paper discusses a number of issues related to the formation of the initial information array, represented by the content of the combined databases, as a carrier of information. The methodological approach to combining information resources in a distributed information processing system, in contrast to the well-known methods of integrating information resources, is based on the procedure for transferring

the technical characteristics data with mismatched attribute values of electronic components described in a series of new national standards. A clarifying action when combining information arrays is the procedure for transferring the technical characteristics data with mismatched values of the CI attribute. This procedure determines a complete set of technical characteristics of the resulting resource, and the method algorithm is a new mechanism compared to the existing and used combining data methods, that described in the national GOST RF standards of the characterics choice for the russian electronic components. Thus, a feature of the developed method is the algorithm for combining industry data and creating a uniform representation of the complete eletronic components data within one form. The article describes and analyzes the process of combining data parts of two information systems. UML activity diagrams and logical conditions for choosing alternative procedures in this algorithm have been developed. Human participation need cases in the process of combining data are given. The use of this approach and the regulatory framework in the form of a series of national GOST R 59988.XX.XX.XX-2022 will allow a standardization of the information resources of the domestic electronic component base and designing the documentation for prototypes of electronic equipment at all stages of the life cycle.
Keywords: information resources, database, information pooling, data pooling, unified information space, data pooling method, analytical computing, information hypercube, decision support system.
For citation: CHUPRINOV, Anatoliy A. Methodological approach to the integration of information resources in a distributed information processing system. IT Security (Russia), [S.l.], v. 30, no. 1, p. 108–122, 2023. ISSN 2074-7136. URL: https://bit.spels.ru/index.php/bit/article/view/1483. DOI: http://dx.doi.org/10.26583/bit.2023.1.08.

Введение

Как показывает практика, для решения задач управления и поддержки принятия решений аналитиками радиоэлектронной отрасли (РЭО) проводится большой объем работ по созданию информационных ресурсов, являющихся входными данными для систем управления и систем поддержки принятия решений (СППР). В настоящее время данная задача усложнилась многократно, т.к. увеличение объемов обрабатываемой информации для РЭО приняло лавинообразный характер. Данный процесс получил название «информационного взрыва» или «информационной перегрузки» [1].

Процесс «информационной перегрузки» в РЭО заключается в том, что многократно возрастает объем информационных потоков при невозможности их оперативной обработки из-за отсутствия специализированных высокоэффективных методов и подходов сбора, хранения и обработки данных в современной сетевой вычислительной среде отрасли. Другими словами, наблюдается противоречие между лавинообразным ростом объемов информации по электронной компонентной базе (ЭКБ), радиоэлектронной аппаратуре (РЭА) и возможностями существующего методического аппарата по сбору, обработке и предоставлению потребителю требуемой информации в заданном виде и в заданное время.

Учитывая то, что каждая аналитическая задача требует «свою» особенную структуру представления данных, необходим механизм (методический подход), позволяющий найти и сформировать в структурированном виде нужную информацию в распределенной информационной среде отрасли и представить ее для решения стоящей перед аналитиком задачи.

1. Метод объединения информации

Процесс формирования доверенного объединенного информационного пространства (ОИП) в распределенной информационной среде РЭО приведен на рис. 1.

Основу методического подхода (далее – Метод) составляют математические модели и методика оценки характеристик доверенной программно-аппаратной среды

функционирования ОИП в условиях информационных угроз на основе метода косвенных признаков [2, 3].

Метод предназначен для формирования начального информационного массива, представленного содержанием объединенных баз данных (БД), как носителя рассматриваемой информации. По этой причине он не затрагивает создание витрин данных и входной информации для алгоритмов и аналитических задач управления и СППР.

Метод объединения информационных ресурсов в системе распределенной обработки информации, в отличие от существующих методов интеграции формационных ресурсов, основан на процедуре переноса данных по техническим характеристикам (ТХ) с несовпадающими значениями атрибутов электронных компонентов.

Начальными условиями данного метода являются информационные ресурсы, представленные в БД, информационными массивами (структурами, массивами, неструктурированными данными и т.д.). Результатом объединения должен стать информационный ресурс, корректно включающий в себя информацию из объединяемых информационных структур (ИС), отвечающий требованиям нормализации, актуальности, однозначность и непротиворечивости.

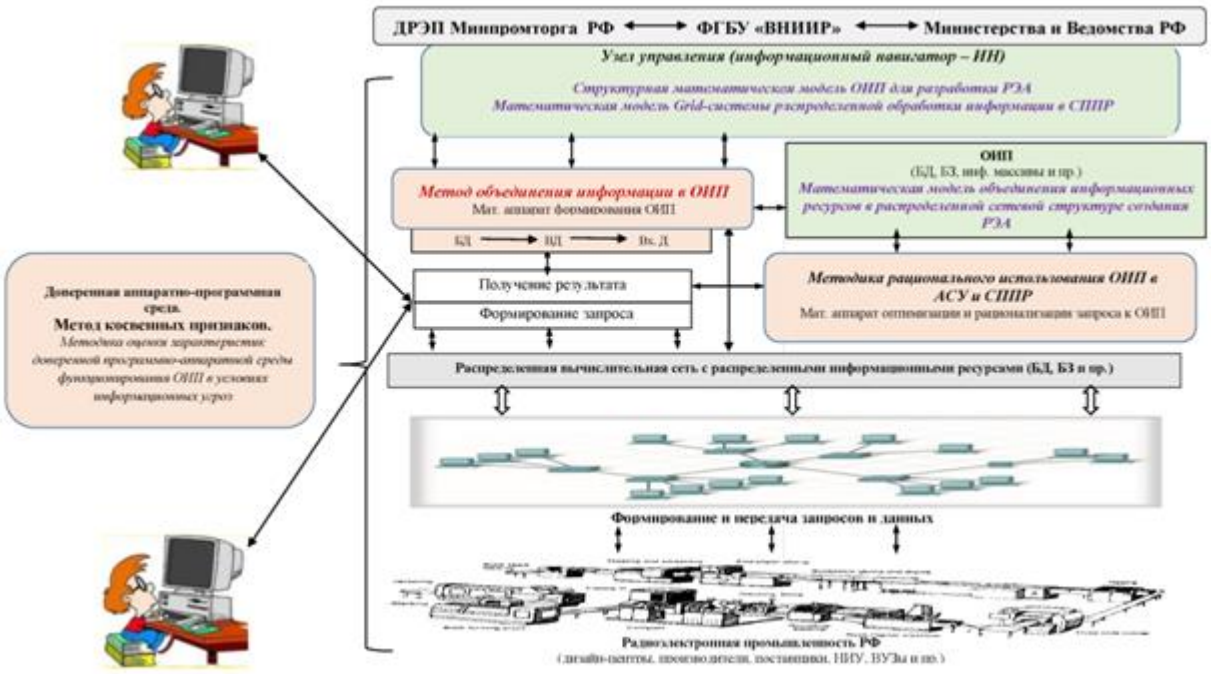

Рис. 1. Методические рекомендации по созданию доверенного объединенного информационного пространства
Fig. 1. Methodological recommendations for creating a trusted unified information space

Сформулируем задачу метода объединения информации в ОИП.

Пусть существуют массивы БД *A* и *B*, относящиеся к одной предметной области (ПрО) знаний, содержащие данные, описывающие объекты, используемые для решения аналитических, поисковых задач и задач проектирования опытных образцов оборудования и техники в данной ПрО.

Необходимо получить интегрированный массив данных в виде БД $C = A \cup B$.

Входными данными для *Метода* являются структурные схемы БД интегрируемых ИС.

Выходными данными – объединенная структурная схема ОИП, которая описывает структуру ИС в рамках их ПрО:

$$SA = \{TB1, TBi, \dots, TBq\}, SB = [TB1, TBi, \dots, TBz), \quad (1)$$

где: *SA и SB – схемы БД, TBi - i-ая таблица в схеме.*

$$TBi = \{ATl, ATJ, \dots, ATm], \quad (2)$$

где: *ATj – атрибуты таблицы; SA и SB принадлежат ИС-1 и ИС-2, поддерживающих БД A и B, соответственно.*

Общая схема *Метода* приведена на рис. 2.

В самом простом варианте, БД – это матрицы заданных размерностей.

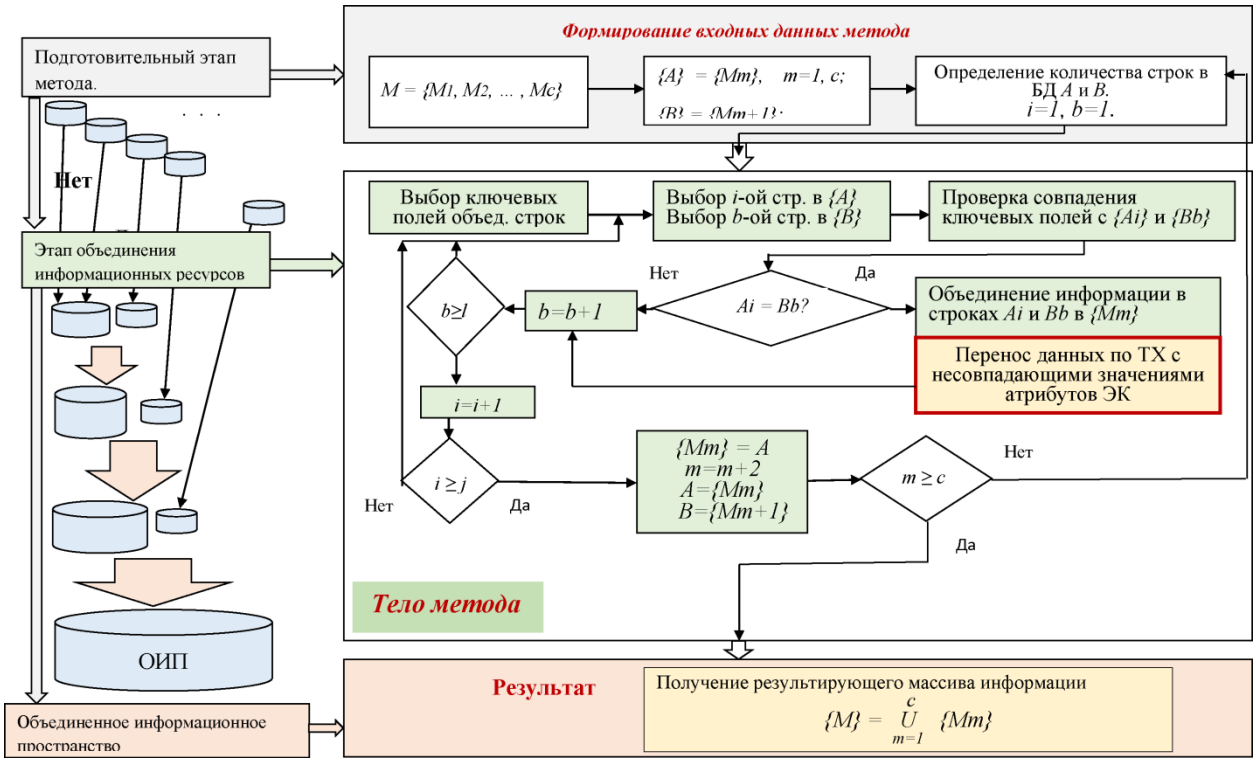

Рис. 2. Общая схема метода объединения информационных ресурсов
Fig. 2. General scheme of the method of combining information resources

Пусть массивы БД *A(d,q)* и *B(r,p)* – матрицы разных размерностей, содержащие в столбцах признаки описания характеристик рассматриваемых объектов, а в строках – описание значений характеристик конкретного объекта где:

$$d = 1, j; \quad q = 1, n; \quad r = 1, l; \quad p = 1, k; \quad m \neq r; \quad q \neq p.$$

Тогда

$$A = \begin{pmatrix} a11 & a12 & a13 & \dots & a1n \\ a21 & a22 & a23 & \dots & a2n \\ a31 & a32 & a33 & \dots & a3n \\ & & \dots & & \\ aj1 & aj2 & aj3 & \dots & ajn \\ & & q & & \end{pmatrix} \downarrow d \qquad B = \begin{pmatrix} b11 & b12 & b13 & \dots & b1k \\ b21 & b22 & b23 & \dots & b2k \\ b31 & b32 & b33 & \dots & b3k \\ & & \dots & & \\ bl1 & bl2 & bl3 & \dots & blk \\ & & p & & \end{pmatrix} \downarrow r$$

$\rightarrow$ $\rightarrow$

Значения характеристик столбцов матриц A и B могут совпадать как по синтаксису, так и по семантике. Тогда эти поля назовем ключевыми. Например, (a'03) – условное обозначение элемента a23 = 106ЛД1А (микросхема), b'01 – условное обозначение элемента – b12 = 106ЛД1А (микросхема). Нулевая строка – название характеристик столбцов матрицы A (атрибут таблицы).

Для однозначной идентификации объектов в матрицах выбирают несколько ключевых полей (атрибутов) и их совпадение в матрицах *A* и *B* позволяет идентифицировать совпадающие строки, описывающие один и тот же объект.

2. Математический аппарат

Рассмотрим математический аппарат, реализующий процедуру объединения матриц *A* и *B,* с учетом особенностей представления технических характеристик ЭКБ ОП, описанных в национальных стандартах.

При автоматизированном анализе значений атрибутов, интегрируемых ИС, делается допущение, что все текстовые описания элементов в той или иной степени несут в себе некоторую семантическую (смысловую) составляющую.

Таким образом, задача заключается в вычислении нескольких примитивных функций, определяющих схожесть атрибутов. Пусть существуют схемы *SA* и *SB:*

$$SA = \{AT1, ATi, \dots, ATn\} \text{ и } SB = \{ABi, ABj, ABm\}, \tag{3}$$

где: *ATi* и *ABj* – атрибуты схемы.

Обнаружение схожести схем заключается в получении коэффициента схожести строк каждого атрибута по функции сравнения строк *f(a↔b)* [4]:

$$\begin{aligned} &f(AT1 \leftrightarrow AB1) = pat1ab1 \\ &f(AT1 \leftrightarrow AB2) = pat1ab2 \\ &f(AT2 \leftrightarrow AB1) = pat2ab1 \\ &f(AT2 \leftrightarrow AB2) = pat2ab2 \\ &\dots \end{aligned} \tag{4}$$

Для каждого вида атрибутов и признаков, по которым их оценивают, используется отдельная функция. Сравнение необходимо проводить, как по признакам с небольшим весом, таким как совпадение типов данных у двух атрибутов, так и по более важным признакам. Например, при анализе структурных элементов, необходимо анализировать названия таблиц и полей из двух интегрируемых схем. При анализе же хранимых ресурсов, следует сравнивать сами данные. Иными словами, возникает задача сравнения строк. Для ее решения предлагается использовать метрики схожести строк [5]. В общем случае, зависимость метрик схожести имеет следующий вид:

$$p = \frac{2xn(X \cap Y)}{n(X)+n(Y)}, \tag{5}$$

где: p – коэффициент похожести строк, $0 \le p \le 1$; X и Y – сравниваемые строки; $n(a)$ – функция, вычисляющая количество символов в строке a.

Однако в таком виде коэффициент p не целесообразно применять для сравнивания строк при большой длине исходных данных и ограниченной длины алфавита. В [4] предлагается расширить данный коэффициент, при помощи использования строк символов – N-граммов. Такой подход позволит оперировать уже не отдельными символами в строке, а набором символов, что позволит находить совпадения наиболее точно.

Пусть существует некоторый алфавит:

$$AB = \{l1, \dots, li\} \tag{6}$$

где: li – отдельный символ алфавита.

В таком случае языком $L(AB)$ на алфавите AB называется множество цепочек конечной длины, состоящих из символов алфавита AB. Высказыванием называется отдельная цепочка на языке $L(AB)$, а N-граммом в алфавите AB называют цепочку символов длинной n. N-грамм может совпадать с некоторым высказыванием, входить в него или не входить вообще в язык. При разбиении строки, необходимо использовать, так называемый, метод шинглов [6], который позволит создать пересекающиеся N-граммы.

Таким образом, используя N-граммы вместо символов в коэффициенте метрики схожести строк, можно привести коэффициент p к следующему виду:

$$p = \frac{2xNgrams(X) \cap Ngrams(Y)}{Ngrams(X) + Ngrams(Y)}\ , \tag{7}$$

где: p – коэффициент похожести строк, $0 \le p \le 1$; X и Y – сравниваемые строки; $Ngrams(a)$ – N-граммы функция, вычисляющая длину N-грамм цепочки от строки X.

Использование не отдельных символов, а целых словосочетаний, позволяет уменьшить количество ложных срабатываний при анализе сходных текстовых строк.

Разработанный подход к сравнению символьных строк предлагается использовать, как для сравнения названий таблиц и их полей, так и данных, хранящихся в интегрируемых БД. Для нахождения полной идентификации сравниваемых объектов используют весовые коэффициенты, позволяющие повысить точность идентификации. С этой целью формируют набор коэффициентов для различных атрибутов по некоторым признакам (по типу данных, по использованию опорных слов, по аннотациям и описаниям и т.д.), обладающим различными весами w:

$$sim(SA,SB) = \{p1w1, pmwm, \dots pkwk\}, \tag{8}$$

где: $sim(SA,SB)$ – *соответствие схем SA и SB;* pm – коэффициент соответствия по некоторому признаку m; wm – вес признака.

Веса для каждого признака определяются вручную, либо при помощи метода машинного обучения [7]. Для определения общей вероятности соответствия элемента схемы БД по совокупности соответствий атрибутов, предлагается использовать линейную функцию:

$$sim(SA,SB) = p1*w1 + p2*w2 + \dots + pn*wn \tag{9}$$

Рассматривая коэффициент соответствия pn, как величину совпадения идентификационных признаков ключевых параметров атрибутов, при которых в соответствие входным атрибутам схемы SA, ставится соответствие выходных атрибутов схемы SB с весом wn, можно рассматривать математическую модель функциональной связи между входными и выходными элементами цепочки символов в виде уравнения регрессии:

$$ySA = f(SB), \tag{10}$$

где: *f(SB)* – некоторая аналитическая зависимость, в качестве которой наиболее часто применяют степенные полиномы:

$$ySA = p0 + p1Sb + p2Sb^2 + \cdots + pmSb^m \tag{11}$$

Полученная зависимость позволяет выделить ключевые атрибуты и соответствующие им весовые характеристики для формирования интегрированных информационных массивов с равновесными значениями атрибутов, что реализует следующие алгоритмы метода.

Для начального объединения двух массивов БД *A* и *B* проведем сравнение и выбор семантически однозначных ключевых полей БД *A* и *B*:

$$\{C'oh=1,k\} = \{a'0i=1,n \cup b'0j=1,k\} = p = \frac{2xNgrams(a0,i) \cap Ngrams(d0,j)}{Ngrams(f0,i)+Ngrams(b0,j)} = \begin{cases}1\\0\end{cases} \tag{12}$$

где: *массив C'0h* – содержит перечень столбцов ключевых полей матриц *A* и *B*, для которых есть семантическое совпадение *Ngrams* строк и *p = 1*.

Объединение матриц происходит последовательным перебором матрицы с большим количеством строк относительно матрицы с меньшим их количеством.

Пусть k > n и j > l; a'01 = b'02, a'03 = b'01, a'0n=b'03.

Размерность результирующей матрицы *C может быть в пределах:* количество строк от *j* до *j*+l; количество столбцов от *k* до *k+n*

$$C = \begin{pmatrix} a'01 & a'02 & a'03 & \ldots & a'0n \\ a11 & a12 & a13 & \ldots & a1n \\ a21 & a22 & a23 & \ldots & a2n \\ a31 & a32 & a33 & \ldots & a3n \\ & \vdots & & & \\ aj1 & aj2 & aj3 & \ldots & ajn \end{pmatrix} \cup \begin{pmatrix} b'01 & b'02 & b'03 & \ldots & b'0k \\ b11 & b12 & b13 & \ldots & b1k \\ b21 & b22 & b23 & \ldots & b2k \\ b31 & b32 & b33 & \ldots & b3k \\ & \vdots & & & \\ bl1 & bl2 & bl3 & \ldots & blk \end{pmatrix} =$$

$$= \begin{pmatrix} \textbf{p11=0 (a01,b01)} & \textbf{p12=0 (a02, b02)} & \textbf{p13=0 (a03, b03)} & \ldots & \textbf{p1n=0 (-, b0k)} \\ \begin{bmatrix} a11 \\ b11 \end{bmatrix} & \begin{bmatrix} a12 \\ b12 \end{bmatrix} & \begin{bmatrix} a13 \\ b13 \end{bmatrix} & & \begin{bmatrix} \\ b1k \end{bmatrix} \\ p21=1(a01) & p22=0(a02, b02) & p23=1(a03) & \ldots & p2n=1(a0n) \\ \begin{bmatrix} a21 \end{bmatrix} & \begin{bmatrix} a22, b22 \end{bmatrix} & \begin{bmatrix} a23 \end{bmatrix} & \ldots & \begin{bmatrix} b2k \end{bmatrix} \\ p31=0 (a01, b0) & p32=0 (a02,b02) & p33=0 (a03,b03) & \ldots & p3n=0 (-, b0k) \\ \begin{bmatrix} a11 \\ b11 \end{bmatrix} & \begin{bmatrix} a12 \\ b12 \end{bmatrix} & \begin{bmatrix} a13 \\ b13 \end{bmatrix} & & \begin{bmatrix} \\ b1k \end{bmatrix} \\ p21=1(a01) & p22=0(a02, b02) & p23=1(a03) & \ldots & p2n=1(a0n) \\ \begin{bmatrix} aj1 \end{bmatrix} & \begin{bmatrix} aj2, bl2 \end{bmatrix} & \begin{bmatrix} aj3 \end{bmatrix} & \ldots & \begin{bmatrix} bjk \end{bmatrix} \end{pmatrix}$$

Результирующая матрица может быть разряженной, т.е. значение ее элементов может быть не определено или равняться пробелу, при использовании текстовых переменных и *0* – при числовых.

Все совпадающие по ключевым полям множества строки объединяются, при этом параметры строки записываются без повторений, что приводит к расширению матрицы от *n* до *k* столбцов, в случае *n < k* или от *k* до *n* в обратном случае. Это действие приводит к увеличению количества столбцов объединенной матрицы до размера матрицы с большим размером столбцов, при этом *n* становится равным *k*.

Все строки, для которых *p=0* добавляются в матрицу с меньшим количеством строк, что приводит к получению результирующей матрицы, равной по размерности матрице с большим количеством строк.

Следующим шагом метода объединения информационных ресурсов в системе распределенной обработки информации является использование механизма MDM для построения ОИП. Объединение информационных ресурсов в общем случае может быть представлено следующей методикой объединения множеств (БД), описанной алгоритмом, приведенным на UML-диаграмме действий, рис. 3.

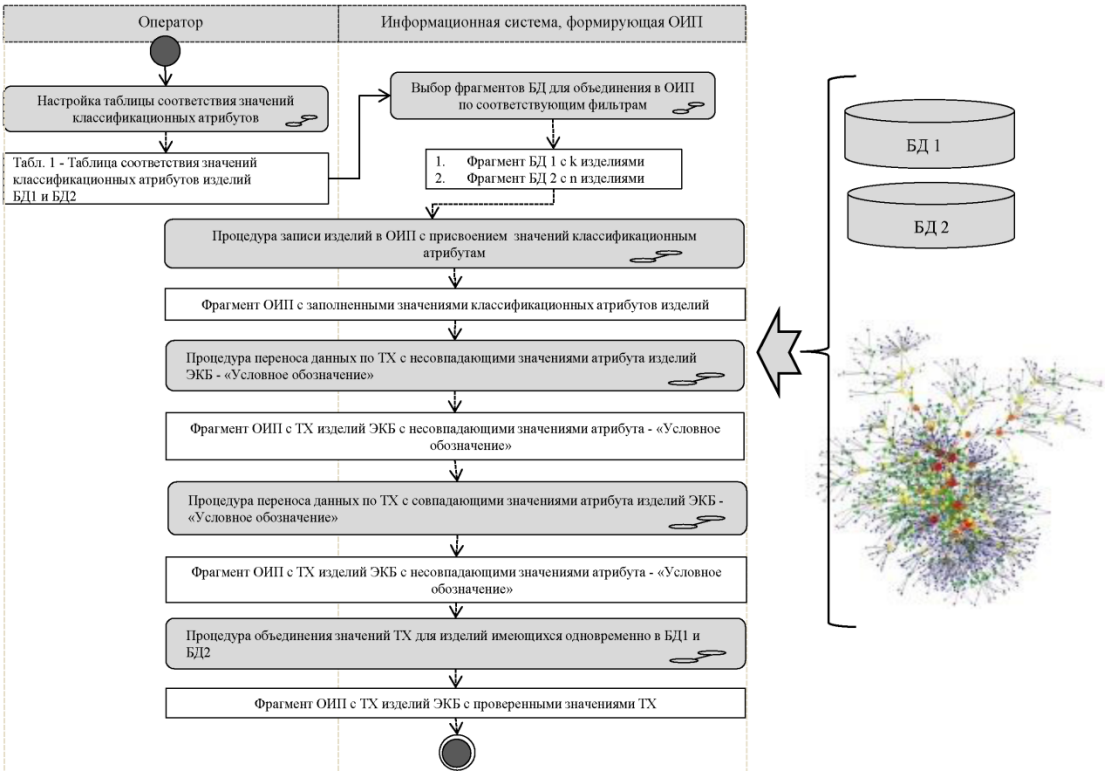

Рис. 3. UML-диаграмма деятельности. Алгоритм методики объединения массивов данных
Fig. 3. UML activity diagram. Algorithm of the method of combining data arrays

Уточняющим действием при объединении информационных массивов является процедура переноса данных по техническим характеристикам (ТХ) с несовпадающими значениями атрибута электронных компонентов (рис. 4). Данная процедура формирует полное множество технических характеристик результирующего ресурса, а алгоритм в методе является новым механизмом на множестве существующих и используемых методов объединения данных [8], описанных в национальных стандартах ГОСТ РФ по выбору ТХ для ЭКБ ОП [9–11].

3. Анализ результатов

Суть метода объединения данных из двух информационных систем заключается в использовании семантического подхода к моделированию данных, опирающегося на онтологическое описание ПрО. Этот подход позволил сформулировать удовлетворительный ответ на вопрос об отношениях сущностей (объектов) и атрибутов ПрО в терминах принадлежности объекта к классу объектов, обладающих данными атрибутами [12]. Существенная черта онтологической модели – это отказ от единого

иерархического дерева и допущение множественности классификаций, в том числе равноправных классификаций объектов и атрибутов.

На рис. 5 представлены фрагменты классификаторов БД1 и БД2 с одинаковыми признаками классификации, но с разной её глубиной. Использование для классификации ЭКБ с одинаковыми классификационными признаками позволяет заранее установить однозначное соответствие классификационных группировок БД1 и БД2 [9]. Это позволяет при объединении данных присвоить значения атрибутам ЭКБ, определяющим классификационную группировку по классификаторам БД1 и БД2 (табл. 1 и 2).

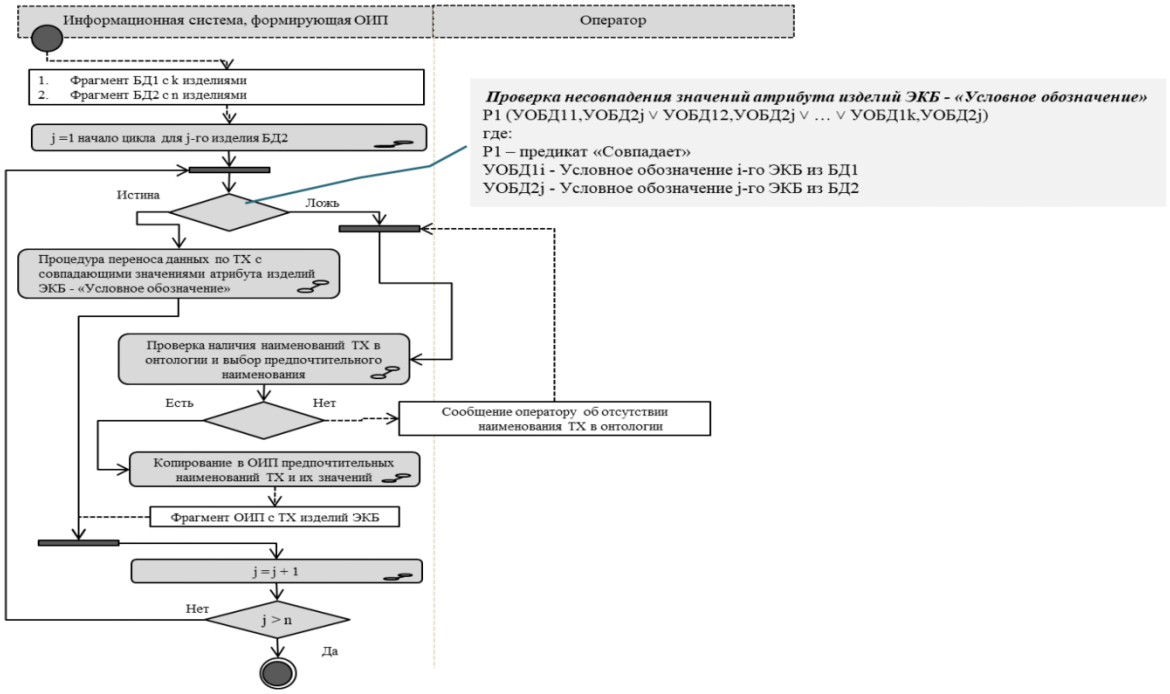

Рис. 4. UML-диаграмма деятельности. Процедура переноса данных по ТХ с несовпадающими значениями атрибута электронных компонентов
Fig. 4. UML activity diagram. Procedure for transferring data on technical characteristics with mismatched attribute values of electronic components

Таблица 1. Соответствия значений классификационных атрибутов для рис. 4

Классификационный атрибут БД1	**Классификационный атрибут БД2**
2 Микросхемы интегральные	2 Микросхемы интегральные
2.2 Микросхемы аналоговые	2.2 Микросхемы аналоговые
2.2.8 Микросхемы аналоговые прочие	2.2.8 Микросхемы аналоговые прочие 2.2.8.1 Аналоговые перемножители сигналов 2.2.8.2 Аттенюаторы 2.2.8.3 Фазовращатели 2.2.8.4 Преобразователи напряжения 2.2.8.5 Микросхемы радиоприемника для применения в устройствах систем связи

Таблица 2. Соответствия значений классификационных атрибутов для рис. 5

Тип изделия	Классификационный атрибут БД 1	Классификационный атрибут БД 2
122УД1А	2.2.1.1 Усилители низкой частоты	***не определено***
122УН2Б	2.2.1.1 Усилители низкой частоты	2.2.1.2 Малошумящие усилители
140УД5АС1ВК	2.2.1.2 Усилители промежуточной частоты	***не определено***
ОС 122УД1А	2.2.1.3 Усилители высокой частоты	***не определено***
140УД6АВК	***не определено***	2.2.1.3 Усилители распределенного усиления
ОС 122УН1А	***не определено***	2.2.1.4 Усилители буферные

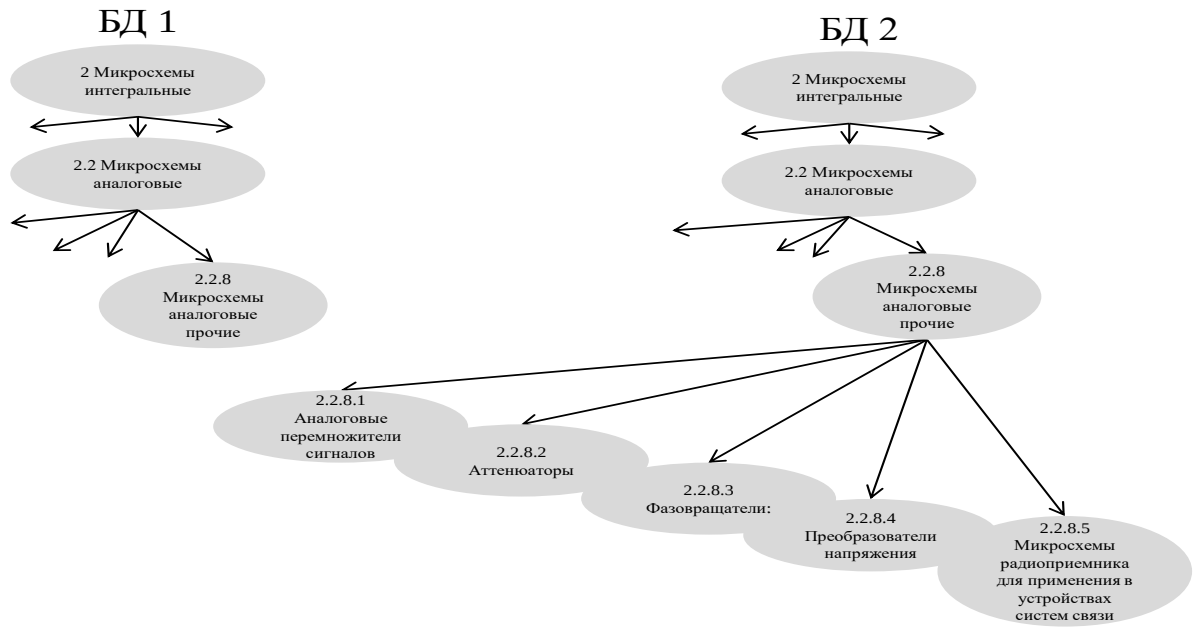

Рис. 5. Фрагменты классификаторов с одинаковыми признаками классификации
Fig. 5. Fragments of classifiers with identical classification features

Из табл. 2 видно, что при объединении данных часть значений классификационных атрибутов ЭКБ будет оставаться неопределенной из-за невозможности формирования однозначной таблицы соответствия значений классификационных атрибутов при использовании разных признаков классификации.

Процедура объединения данных по ТХ с несовпадающими значениями атрибута тип ЭКБ – «Условное обозначение» сводится к копированию наименований ТХ и их значений из БД в формируемый фрагмент ОИП.

На рис. 6 представлены фрагменты классификаторов БД1 и БД2 с разными признаками классификации. В БД1 классификационный признак – рабочий диапазон частот усилителя, в БД2 – его функциональное назначение.

Отметим, что для идентификации изделий ЭКБ используется атрибут – «Условное обозначение», являющийся уникальным для каждого изделия.

Подробнее рассмотрим более сложную процедуру объединения данных для изделия 122УН2Б в БД1 и БД2 (рис. 5), по которой необходимо объединить данные этих ИС. На примере этого изделия опишем алгоритм метода объединения данных по ТХ [11].

В табл. 3 представлены ТХ рассматриваемого изделия, имеющиеся (предположительно) в БД1 и БД2, с уникальными номерами (УН ТХ), которые присвоены в серии национальных стандартов [9, 10, 11, 13].

Из табл. 3 видно:

– ТХ с порядковыми номерами 1,…,6 имеются в БД1 и отсутствуют в БД2;
– ТХ с порядковыми номерами 7,…,9 имеются в БД2 и отсутствуют в БД1.

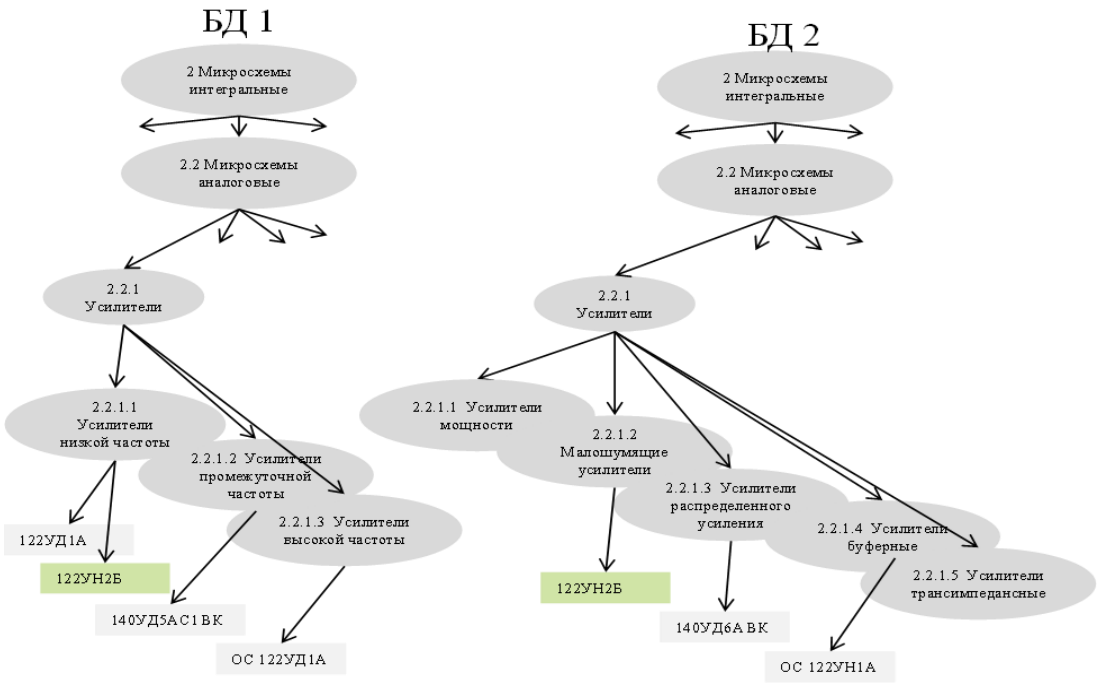

Рис. 6. Фрагменты классификаторов с разными признаками классификации
Fig. 6. Fragments of classifiers with different classification features

Так как данные по рассматриваемым изделиям объединяются, то в формируемом фрагменте ОИП должны присутствовать все вышеупомянутые ТХ с порядковыми номерами 1, …,9. В случае, если по какой-либо ТХ и в какой-либо БД будет отсутствовать значение атрибута УН ТХ, то наличие (отсутствие) этой ТХ в БД проверяется по её наименованию с использованием онтологии данной области [9, 10, 11] в которой имеются синонимичные наименования этой ТХ и УН ТХ.

Используя исчисление предикатов первого порядка [14], представим алгоритм объединения данных по ТХ с порядковыми номерами 10, …,12 в формируемый фрагмент ОИП, имеющийся и в БД1 и в БД2, но с разными наименованиями (табл. 3):

Шаг 1. Определяем первую строку записи в БД1.

Цикл по *j=1 до j=**kon**, где*: ***kon***-количество записей в БД1.

Используя классификационные атрибуты изделия 122УН2Б, выбираем фрагмент онтологии из [13, 15], к которому оно относится.

Шаг 2. Проводим поиск ТХ из БД1 в выбранном фрагменте онтологии ПрО. Поиск может осуществляться как по УН ТХ из БД1, так и по наименованию ТХ из БД1.

Шаг 2.1. Поиск по множеству УН ТХ онтологии осуществляется пока предикат не примет значение «истина» или не переберет все УН ТХ в выбранном фрагменте онтологии:

Цикл от i = 1 до i = kyn

$\mathbf{P(yn1, ynoi)}$, где: ***kyn*** – количество УН ТХ в выбранном фрагменте онтологии;

P – предикат «Совпадает»;
yn1 – переменная с УН ТХ из БД1;
ynoi – переменная с i-м УН ТХ из онтологии ПрО.

Шаг 2.2. Поиск по множеству наименований ТХ онтологии осуществляется пока предикат не примет значение «истина» или не переберет все наименования ТХ в выбранном фрагменте онтологии:

Цикл от i = 1 до i = kn
P(n1, noi) , где: ***kn*** – количество наименований ТХ в выбранном фрагменте онтологии;
P – предикат «Совпадает»;
n1 – переменная с наименованием ТХ из БД1;
noi – переменная с i-м наименованием ТХ из онтологии ПрО.

Шаг 3. В случае наличия предиката со значением «истина» присваиваем ТХ предпочтительное (эталонное) наименование из онтологии и выполняем соответствующую запись в ОИП.

В табл. 4 приведен фрагмент онтологии, по которому видно, что эталонное наименование 11-й ТХ с УН 2.4.3 – это «Сопротивление в открытом состоянии» (выделено жирным шрифтом), именно его переносим в ОИП.

Иначе, формируем сообщение оператору системы, что ТХ с данным наименованием в онтологии отсутствует. Переходим к **Шагу 1**.

По вышеописанному алгоритму обрабатываем ТХ с порядковыми номерами 10, …,12 из БД1 и из БД2. В результате из 9 ТХ БД1 и 6 ТХ БД2 получаем фрагмент ОИП из 12 ТХ изделия 122УН2Б (табл. 5). Для ОИП выбираются наименования ТХ, отмеченные в [13] как предпочтительные (эталонные).

Таблица 3. ТХ изделия 122УН2Б по [13, 15]

№ п/п	УН ТХ	Наименование ТХ в БД 1	Наименование ТХ в БД 2
1	4.1	Тип корпуса	-
2	2.1.11	Напряжение питания	-
3	2.2.11	Ток потребления	-
4	3.1	Рабочая температура	-
5	2.1.9	Максимальное выходное напряжение	-
6	2.1.15	Напряжение смещения нуля	-
7	2.2.17	-	Ток утечки
8	1.1.9	-	Время включения
9	1.1.10	-	Время выключения
10	2.1.18	Падение напряжения на интегральной микросхеме	Падение напряжения
11	2.4.3	Сопротивление открытого ключа	Сопротивление в открытом состоянии
12	1.3.17	Скорость нарастания выходного напряжения максимальная	Максимальная скорость нарастания выходного напряжения

Таблица 4. Фрагмент онтологии из [11]

УН ТХ	**Наименование ТХ**	**Описание (физический смысл ТХ)**
2.4.3	**Сопротивление в открытом состоянии** по ГОСТ 57441-2017 (п. 2, № 75), ГОСТ 19480-89 (табл. 1, п. 140), ОСТ 11 0278-87 (с. 17) Синонимы: – Сопротивление в открытом состоянии интегральной микросхемы по ГОСТ 19480-89 (табл. 1, п. 140), ОСТ 11 0278-87 (с. 17) – Сопротивление открытого ключа (технические условия – ТУ)	1. Сопротивление в открытом состоянии – отношение падения напряжения между входом и соответствующим выходом микросхемы к току, протекающему через этот выход, в заданном режиме. По ГОСТ 57441-2017 (п. 2, № 75). 2. Отношение падения напряжения между аналоговым входом к вызвавшему его току при открытом (включенном) канале по ОСТ 11 0278-87 с. 17).

Остается провести операцию по объединению значений ТХ с порядковыми номерами 10, …,12 для включения в формируемый фрагмент ОИП, которые, в случае отсутствия ошибок, должны быть одинаковыми. Операция представляет собой проверку вышеуказанного утверждения и, если это не так, то необходимо пометить эти значения, как ошибочные, и дать соответствующий комментарий для оператора, который по ТУ на изделия 122УН2Б проверит и исправит ошибочно введенные значения.

Таблица 5. Фрагмент ОИП по ТХ изделия 122УН2Б

№ п/п	**УН ТХ**	**Наименование ТХ в ОИП**
1	4.1	Тип корпуса
2	2.1.11	Напряжение питания
3	2.2.11	Ток потребления
4	3.1	Рабочая температура
5	2.1.9	Максимальное выходное напряжение
6	2.1.15	Напряжение смещения нуля
7	2.2.17	Ток утечки
8	1.1.9	Время включения
9	1.1.10	Время выключения
10	2.1.18	Падение напряжения
11	2.4.3	Сопротивление в открытом состоянии
12	1.3.17	Максимальная скорость нарастания выходного напряжения

Заключение

Применение разработанного метода объединения информационных множеств позволяет решить задачу формирования начального вида гиперкуба информации, необходимого для аналитических вычислений, в частности, для обработки в системах поддержки принятия решений при создании образцов РЭА в доверенной вычислительной распределенной среде разработки с использованием ЭКБ отечественного производства.

СПИСОК ЛИТЕРАТУРЫ:

1. Петрович Н.Т. Люди и биты. Информационный взрыв: что он несет. М.: Знание. 1986. – 110 с.
2. Чуприков Анатолий А. Математическая модель формирования объединенного информационного пространства радиоэлектронной отрасли. Безопасность информационных технологий, [S.l.], т. 28, № 4, с. 127–138, 2021. ISSN 2074-7136. DOI: http://dx.doi.org/10.26583/bit.2021.4.10. – EDN: ZNKHZN.
3. Чуприков, Анатолий А.; Смирнов, Дмитрий О. Метод косвенных признаков для выявления аппаратных угроз технических средств. Безопасность информационных технологий, [S.l.], т. 29, № 2, с. 10–19, 2022. ISSN 2074-7136. DOI: http://dx.doi.org/10.26583/bit.2022.2.01. – EDN: WRDVOJ.
4. Семерханов И.А. Методы и алгоритмы автоматизированной интеграции информационных ресурсов на основе онтологического подхода: - Диссертация кандидата технических наук: 05.13.12. Семерханов Илья Александрович. Санкт-Петербург, 2014. – 140 с.
5. Матье Констан. Сходство между словами, магистр компьютерных наук, Парижский университет в Марн-ла-Валле, 2012.
6. Сотников И.М. Применение «Алгоритма шинглов» при обработке заданий на курсовое проектирование. Современные проблемы науки и образования. 2014, № 4. URL: https://science-education.ru/ru/article/view?id=14452 (дата обращения: 02.02.2023).
7. Дж. Вандер Плас. Python для сложных задач. Наука о данных и машинное обучение. ISBN: 978-5-496-03068-7. СПб.: Питер, 2018. – 576 с.
8. Климов С.М., Чуприков А.А. Алгоритм объединения данных по электронным компонентам из двух информационных систем. Сб. трудов 4 ЦНИИ МО РФ, г. Королев. 2022, № 171, т. 2, с. 35–40.
9. Савин М.Л., Корчагин А.И., Чуприков А.А., Колядин А.И., Егоркин А.В. ГОСТ Р РФ 59988.00.0-2022 Системы автоматизированного проектирования электроники. Информационное обеспечение. Технические характеристики электронных компонентов. Общие положения. Стандартинформ, издание официальное. Москва, 2022. – 17 с.
10. Савин М.Л., Корчагин А.И., Чуприков А.А., Колядин А.И., Егоркин А.В. ГОСТ Р 59988.02.2-2022 Системы автоматизированного проектирования электроники. Информационное обеспечение. Микросхемы интегральные. Перечень технических характеристик. Стандартинформ, издание официальное. Москва, 2022. – 47 с.
11. Савин М.Л., Корчагин А.И., Чуприков А.А., Колядин А.И., Егоркин А.В. ГОСТ Р 59988.02.1-2022 Системы автоматизированного проектирования электроники. Информационное обеспечение. Микросхемы интегральные. Спецификации декларативных знаний по техническим характеристикам. Стандартинформ, издание официальное, Москва. 2022. – 97 с.
12. Алексеев В.В., Боков С.И., Колядин А.И., Чуприков А.А. Вопросы организации (структурирования) информации для системы управления знаниями радиоэлектронной отрасли. НАНО-индустрия, спецвыпуск 2020 (89), Международный форум «Микроэлектроника-2019», 5-я международная научная конференция «Электронная компонентная база и микроэлектронные модули», Сборник докладов., Республика Крым, г. Алушта, 01-06 октября 2019 г., Техносфера, (Часть 1) с. 246–256. URL: https://publications.hse.ru/pubs/share/direct/375140834.pdf (дата обращения: 02.02.2023).
13. Савин М.Л., Корчагин А.И., Чуприков А.А., Колядин А.И., Егоркин А.В., Смирнов Н.С. ГОСТ Р 59988.03.1-2022 Системы автоматизированного проектирования электроники. Информационное обеспечение. Приборы и модули полупроводниковые. Спецификации декларативных знаний по техническим характеристикам. Стандартинформ, издание официальное, Москва. 2022. – 74 с.
14. Мендельсон Э. Введение в математическую логику. Пер. с англ. М: Издательство «Наука», 1971. – 320 с.
15. Савин М.Л., Корчагин А.И., Чуприков А.А., Колядин А.И., Егоркин А.В., Смирнов Н.С. ГОСТ Р 59988.03.2-2022 Системы автоматизированного проектирования электроники. Информационное обеспечение. Приборы и модули полупроводниковые. Перечень технических характеристик. Стандартинформ, издание официальное, Москва, 2021. – 107 с.

REFERENCES:

[1] Petrovich N.T. People and bits. Information explosion: what he carries. M.: Znanie. 1986. – 110 p. (in Russian).
[2] Chuprinov Anatoliy A. Mathematical model of the formation of a unified information space of the radio-electronic industry. IT Security (Russia), [S.l.], v. 28, no. 4, p. 127–138, 2021. ISSN 2074-7136. DOI: http://dx.doi.org/10.26583/bit.2021.4.10 (in Russian). – EDN: ZNKHZN.
[3] Chuprinov, Anatoliy A.; Smirnov, Dmitry O. Indirect signs method for detecting hardware threats of technical means. IT Security (Russia), [S.l.], v. 29, no. 2, p. 10–19, 2022. ISSN 2074-7136. DOI: http://dx.doi.org/10.26583/bit.2022.2.01 (in Russian). – EDN: WRDVOJ.

[4] Semerkhanov I.A. Methods and algorithms of automated integration of information resources based on the ontological approach: - Dissertation of Candidate of Technical Sciences: 05.13.12. Semerkhanov Ilya Alexandrovich. St. Petersburg, 2014. – 140 p. (in Russian).
[5] Mathieu Constant. Similarity between words, Master of Computer Science, University of Paris in Marne-la-Vallee, 2012.
[6] Sotnikov I.M. The application of the "Shingle algorithm" in the processing of assignments for course design. Modern problems of science and education. 2014, № 4. URL: https://science-education.ru/ru/article/view?id=14452 (accessed: 02.02.2023) (in Russian).
[7] J. Vander Place. Python for complex tasks. Data Science and Machine learning. ISBN: 978-5-496-03068-7. St. Petersburg: Peter. – 576 p. (in Russian).
[8] Klimov S.M., Chuprinov A.A. Algorithm for combining data on electronic components from two information systems. Proceedings of the 4th Central Research Institute of the Ministry of Defense of the Russian Federation, Korolev. 2022, no. 171, vol. 2, p. 35–40 (in Russian).
[9] Savin M.L., Korchagin A.I., Chuprinov A.A., Kolyadin A.I., Egorkin A.V. GOST R RF 59988.00.0-2022 Systems of computer-aided design of electronics. Information support. Technical characteristics of electronic components. General provisions. Standartinform, official publication. Moscow, 2022. – 17 p. (in Russian).
[10] Savin M.L., Korchagin A.I., Chuprinov A.A., Kolyadin A.I., Egorkin A.V. GOST R 59988.02.2-2022 Systems of computer-aided design of electronics. Information support. Integrated circuits. List of technical characteristics. Standartinform, official publication. Moscow, 2022. – 47 p. (in Russian).
[11] Savin M.L., Korchagin A.I., Chuprinov A.A., Kolyadin A.I., Egorkin A.V. GOST R 59988.02.1-2022 Systems of computer-aided design of electronics. Information support. Integrated circuits. Specifications of declarative knowledge on technical characteristics. Standartinform, official publication, Moscow. 2022. – 97 p. (in Russian).
[12] Alekseev V.V., Bokov S.I., Kolyadin A.I., Chuprinov A.A. Questions of organization (structuring) of information for the knowledge management system of the radioelectronic industry. NANO-industry, special issue 2020 (89), International Forum "Microelectronics-2019", 5th International Scientific Conference "Electronic component base and Microelectronic modules", Collection of reports., Republic of Crimea, Alushta, 01-06 October 2019, Technosphere, (Part 1) p. 246–256. URL: https://publications.hse.ru/pubs/share/direct/375140834.pdf (accessed: 02.02.2023) (in Russian).
[13] Savin M.L., Korchagin A.I., Chuprinov A.A., Kolyadin A.I., Egorkin A.V., Smirnov N.S. GOST R 59988.03.1-2022 Systems of computer-aided design of electronics. Information support. Semiconductor devices and modules. Specifications of declarative knowledge on technical characteristics. Standartinform, official publication, Moscow. 2022. – 74 p. (in Russian).
[14] Mendelssohn E. Introduction to mathematical logic. Trans. from English. M: Publishing House "Science", 1971. – 320 p. (in Russian).
[15] Savin M.L., Korchagin A.I., Chuprinov A.A., Kolyadin A.I., Egorkin A.V., Smirnov N.S. GOST R 59988.03.2-2022 Systems of computer-aided design of electronics. Information support. Semiconductor devices and modules. List of technical characteristics. Standartinform, official publication, Moscow, 2021. – 107 p. (in Russian).

Поступила в редакцию – 8 февраля 2023 г. Окончательный вариант – 17 февраля 2023 г.
Received – February 08, 2023. The final version – February 17, 2023.