

**МЕЖДУНАРОДНЫЙ ЦЕНТР НАУЧНОГО СОТРУДНИЧЕСТВА  
«НАУКА И ПРОСВЕЩЕНИЕ»**



# **НАУЧНОЕ ОБОЗРЕНИЕ: АКТУАЛЬНЫЕ ВОПРОСЫ ТЕОРИИ И ПРАКТИКИ**

**СБОРНИК СТАТЕЙ XXI МЕЖДУНАРОДНОЙ НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ,  
СОСТОЯВШЕЙСЯ 10 ФЕВРАЛЯ 2026 Г. В Г. ПЕНЗА**

**ПЕНЗА  
МЦНС «НАУКА И ПРОСВЕЩЕНИЕ»  
2026**

УДК 001.1  
ББК 60  
НЗ4

Ответственный редактор:  
Гуляев Герман Юрьевич, кандидат экономических наук

НЗ4

**НАУЧНОЕ ОБОЗРЕНИЕ: АКТУАЛЬНЫЕ ВОПРОСЫ ТЕОРИИ И ПРАКТИКИ:**  
сборник статей XXI Международной научно-практической конференции. – Пенза:  
МЦНС «Наука и Просвещение». – 2026. – 284 с.

ISBN 978-5-00268-350-5

Настоящий сборник составлен по материалам XXI Международной научно-практической конференции **«НАУЧНОЕ ОБОЗРЕНИЕ: АКТУАЛЬНЫЕ ВОПРОСЫ ТЕОРИИ И ПРАКТИКИ»**, состоявшейся 10 февраля 2026 г. в г. Пенза. В сборнике научных трудов рассматриваются современные проблемы науки и практики применения результатов научных исследований.

Сборник предназначен для научных работников, преподавателей, аспирантов, магистрантов, студентов с целью использования в научной работе и учебной деятельности.

Ответственность за аутентичность и точность цитат, имен, названий и иных сведений, а также за соблюдение законодательства об интеллектуальной собственности несут авторы публикуемых материалов.

Полные тексты статей в открытом доступе размещены в Научной электронной библиотеке **Elibrary.ru** в соответствии с Договором №1096-04/2016К от 26.04.2016 г.

УДК 001.1  
ББК 60

© МЦНС «Наука и Просвещение» (ИП Гуляев Г.Ю.), 2026  
© Коллектив авторов, 2026

ISBN 978-5-00268-350-5

# ФИЗИКО-МАТЕМАТИЧЕСКИЕ НАУКИ

УДК 004.021

# АЛГОРИТМ ТОЧНОГО СОПОСТАВЛЕНИЯ ДАННЫХ СИСТЕМЫ ФОРМИРОВАНИЯ ОПТИМАЛЬНЫХ ПРЕДЛОЖЕНИЙ НА ВЫБОР ИЗДЕЛИЙ ЭЛЕКТРОННОЙ КОМПОНЕНТНОЙ БАЗЫ В ПРОЦЕССАХ РАЗРАБОТКИ РАДИОЭЛЕКТРОННОЙ АППАРАТУРЫ

РУБЦОВ ЮРИЙ ВАСИЛЬЕВИЧ

генеральный директор  
АО "ЦКБ "Дейтон",

124681, Российская Федерация, г. Москва, г. Зеленоград, корпус 100

**Аннотация:** В статье показаны результаты исследований алгоритмов сопоставления, проведены обобщения, выполнен анализ и разработан алгоритм точного сопоставления для нормализации данных системы формирования оптимальных предложений на выбор изделий электронной компонентной базы в процессах разработки радиоэлектронной аппаратуры.

Проведен анализ результатов исследований и сформулированы требования по подготовке данных к сопоставлению. Определены концепция, правила, идентификаторы, блокирующие переменные и частота обновления данных, необходимые для построения четкой последовательности действий алгоритма, достаточной для создания компьютерных программ, эффективно обеспечивающих функционирование автоматизированной системы.

Новизной полученного решения является практическая реализация алгоритма точного сопоставления, к которому впервые применены представленные в статье разработанные методы, подтвердившие свою результативность в работающей автоматизированной системе Дейтон, результаты функционирования которой используют предприятия радиоэлектронной промышленности.

**Ключевые слова:** алгоритм точного сопоставления данных, машинное обучение, радиоэлектронная аппаратура, электронная компонентная база.

ALGORITHMS FOR PRECISE COMPARISON OF DATA IN THE SYSTEM FOR GENERATING OPTIMAL PROPOSALS FOR SELECTING ELECTRONIC COMPONENTS IN THE PROCESSES OF CONSTRUCTION ELECTRONIC EQUIPMENT

Rubtsov Yuri Vasilievich

**Annotation:** The article presents the results of studies of matching algorithms, generalizations are made, an analysis is performed and an algorithm of exact matching is developed for normalization of data of the system of formation of optimal proposals for selection of products of the electronic component base in processes of development of radio-electronic equipment.

The analysis of results of studies is carried out and requirements for preparation of data for matching are

formed. The concept, rules, identifiers, blocking variables and frequency of data update are defined, necessary for construction of a clear sequence of actions of the algorithm, sufficient for creation of computer programs, effectively ensuring functioning of the automated system.

The novelty of the obtained solution is the practical implementation of the exact matching algorithm, to which the developed methods presented in the article were applied for the first time, which confirmed their effectiveness in the operating automated system Deuteron, the results of the operation of which are used by enterprises of the radio-electronic industry.

**Keywords:** exact data matching algorithm, machine learning, electronic equipment, electronic components

### Введение

Разработка радиоэлектронной аппаратуры (РЭА) является сложным процессом. Особенное место в нем занимает выбор изделий электронной компонентной базы (ЭКБ) для применения РЭА. Информация об ЭКБ для этого должна быть достаточного качества.

Источниками, поставляющими такую информацию являются изготовители и потребители изделий ЭКБ, исследовательские и испытательные организаций. Информация поступает из конструкторских и технологических документов; результатов исследований; измерений; испытаний; применения; эксплуатации; модернизации и утилизации.

Разнообразие источников и форматов информации требует проведения сопоставления данных, выявить избыточность и обеспечить актуальность информации на этапах ее сбора, обобщения и анализа.

Сопоставление данных – это процесс выделения существенных общих свойств в наборах данных для сравнения элементов наборов данных и получения подтверждения их отнесения к одному и тому же изделию ЭКБ.

Сопоставление данных включает:

- выделение существенных свойств данных, общих для целей сопоставления;
- сравнение и анализ данных для установления их идентичности или отличительных черт. В этих целях автором статьи применены результаты исследований технологии ICMH (Information Classification, Marking and Handling) – совокупности взаимосвязанных и взаимодействующих методов и инструментов, применяемых для решения задач сбора, обработки и анализа информации и получения достоверных данных в процессах сбора, обработки и анализа информации [1].

Единицей измерения сопоставляемых данных является набор данных. Наборы данных включают различные типы информации: числовой, текстовый, изображения, видео и аудио, и могут храниться в различных форматах.

Для целей обеспечения работы Системы, наборы данных после сопоставления имеют структурированный вид. Они преобразованы в четко определенный формат. Для предоставления доступа к ним и их обработки.

В сопоставлении данных используется процедура сравнения - анализа различий между наборами данных. Это выполняется для обнаружения ошибок, изменений, дубликатов или для получения понимания структуры и содержания данных.

Для успешного выполнения сопоставления необходимо решить проблему неоднородности данных. Неоднородность данных характеризуется ситуацией, когда часть данных существенным образом отличается от основного массива данных. Подобные особо отличающиеся данные являются выбросами. Как правило, выбросы исключаются из наборов. В основном, это выполняется вручную с помощью специалиста, который приводил данные в вид, необходимый для их дальнейшей обработки. Эта ручная работа утомительна, подвержена ошибкам и малоэффективна.

Неоднородность данных остается постоянной проблемой из-за разнообразия источников, форматов и методологий, используемых для формирования и передачи данных источниками.

Фундаментом для решения этих проблем выступает стандартизация данных. Ее основная цель – обеспечить соответствие различных источников данных общему набору правил и показателей, что поз-

воляет проводить содержательное и точное сопоставление данных за счет сведения к минимуму расхождений. Благодаря применению согласованных определений, единиц измерения и способов представления информации, сопоставление становится более простым, прозрачным и результативным.

Цель показанных в статье исследований – автоматизировать обнаружение сходства между данными, используя алгоритмы в качестве инструмента. Алгоритмы сопоставления принимают исходные данные, вычисляют базовое сходство между ними, сравнивают содержания и типы данных. Затем они вычисляют значения сходства и формируют результаты сопоставления данных. В статье также показаны варианты и сценарии реализации алгоритма.

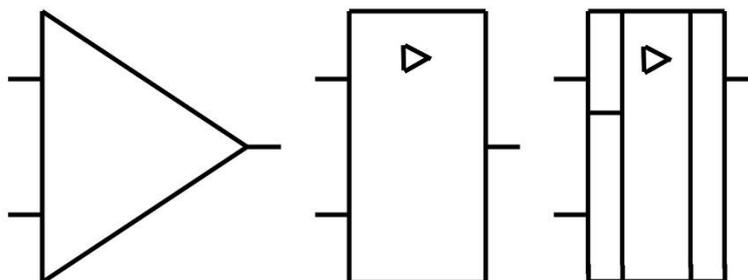
Сопоставление данных – это сложный процесс определения того, какие данные от различных источников соответствуют одному тому же изделию ЭКБ. Это необходимо для обеспечения нормализации данных и оптимального выбора изделий ЭКБ в системе (Система) формирования предложений на применение изделий ЭКБ в радиоэлектронной аппаратуре.

Почему это сложно? Потому что необходимо сопоставить данные из различных источников разных форматов. Они могут отличаться, но содержать одну и ту же информацию. Например:

1) варианты написания, псевдонимы, изменение имен: условное обозначение микроконтроллера, МІК32 АМУР (К1948ВК018) 32-Бит, RISC-V, 32МГц, ОTR 256 бит QFN-64. К1948ВК018 – это установленные регулятором условное обозначение. Может иметь различные варианты написания – латиница и кириллица, указываются через пробел в различных местах К 1948ВК018, К 1948ВК 018 др. Дополнительное условное обозначение (синонимы) МІК32 АМУР могут иметь различные варианты написания. Для информативности источники указывают в условном обозначении параметры изделия: 32-Бит, RISC-V, 32МГц, ОTR 256 бит и даже конструктивное исполнение корпуса микросхемы QFN-64 – (Quad Flat No-leads package) – имеющий планарные 64 вывода, расположенных непосредственно под микросхемой по всем четырём сторонам. Корпус имеет квадратную форму. На самом деле имеются корпуса QFN прямоугольной формы. Особенности таких обозначений исследованы с участием автора в [2];

2) различные варианты написания дат, величин и единиц измерения: 5.05.2025 или 05.05.2025 или 05/05/2025; 100 Вт или 100 W или 0,1 кВт;

3) наличие графических объектов, несущих информацию (условные графические обозначения (УГО), схемы включения, графики зависимостей, габаритные чертежи). На рис. 1 представлены УГО микросхем операционных усилителей в различных вариантах исполнения, сложных для решения задач сопоставления.



**Рис. 1. УГО микросхем операционных усилителей в различных вариантах исполнения**

4) наличие дополнительных информационных материалов. В качестве таковых могут быть модели ЭКБ для применения в системах автоматизированного проектирования РЭА. Их применение существенно повышает эффективность разработки РЭА. А наличие является не обязательным, но дополнительным фактором преимуществ изделия ЭКБ в выборе для применения в РЭА. Необходимость наличия таких моделей в составе информации об ЭКБ их особенности исследованы автором в [3, 4].

После сопоставления таких данных, между ними устанавливаются связи. Идентичные данные могут объединяться или удаляться их дубликаты, для обеспечения согласованности. Этот процесс со-

держит интеграцию данных и включает объединение данных, находящихся в различных источниках, и предоставление данных в Системе в унифицированном виде.

### Результаты исследований и формирование требований по подготовке данных к сопоставлению

Определено, результаты сопоставления зависят от качества данных. Для обеспечения успешного и точного сопоставления, данные проверяются на полноту с четким пониманием охвата каждого элемента из них. Участвующие в сопоставлении данные, могут иметь ошибки. Некоторые обнаруживаются путем проверки логической согласованности данных: например, бескорпусная микросхема не может быть в корпусе. При этом в ней может применяться многовыводная рамка. Результаты сбора, анализа и обобщения информации о многовыводных рамках, как элементов изделий ЭКБ, исследованы и опубликованы автором в [5]. Вторым примером могут быть диапазоны числовых элементов, которые проверяются на принадлежность к изделиям, например, диапазоны температур для изделий конкретной категории качества. Они регламентируются документами по стандартизации для каждого типа изделия.

Элементы наборов данных могут включать в себя объединенные строки, например, функциональное назначение и область применения. Эти элементы разбираются. Пример показан на рис. 2.

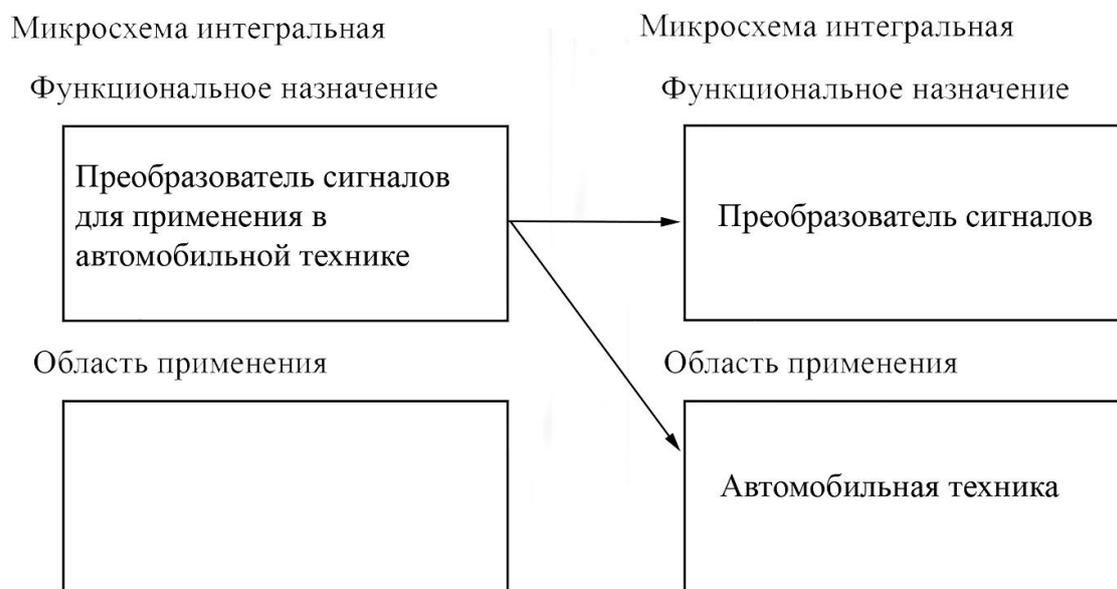
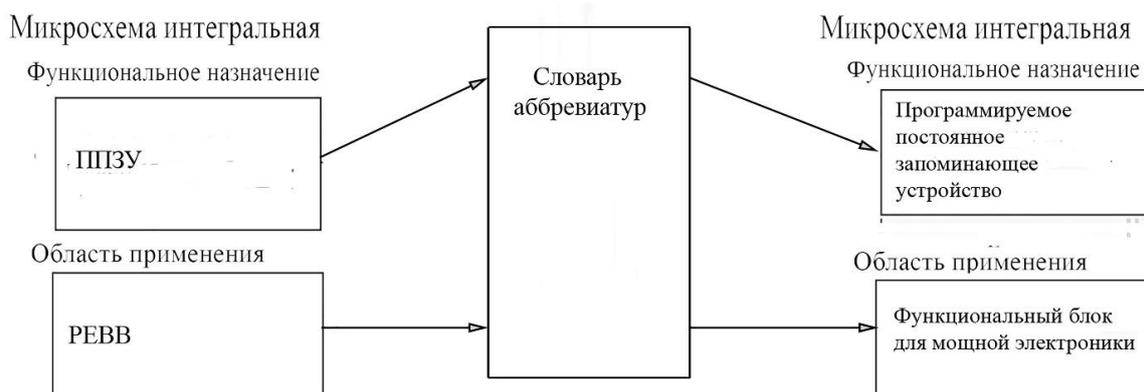


Рис. 2. Пример разбора элементов набора данных

Варианты для разбора могут быть разные. Различия в написании или использование аббревиатур увеличивает сложность сопоставления. Решение реализуется с помощью словарей, которые выравнивают разные варианты написания и могут быть адаптированы к источникам информации. Пример показан на рис. 3.

Другие ошибки, которые рассматриваются и исправляются, – это транслитерации в элементах данных. Для исправления ошибок могут использоваться словари, сформированные на основании ГОСТ 7.79-2000 «Система стандартов по информации, библиотечному и издательскому делу. Правила транслитерации кирилловского письма латинским алфавитом», а также рекомендации ГОСТ Р 7.0.34-2014 «Система стандартов по информации, библиотечному и издательскому делу. Правила упрощенной транслитерации русского письма латинским алфавитом» в обратном порядке.

Строки могут содержать слова плеоназмы (дублирование некоторого смысла), такие как «максимальное граничное напряжение», «предельная минимальная температура», «микросхема электронного устройства триггер». Эти избыточные слова удаляются, путем прямой ссылки на словарь, содержащий список таких избыточных слов.



**Рис. 3. Пример применения словаря аббревиатур**

Отсутствующие значения последовательно помечаются как таковые.

Все данные должны иметь стандартизированные форматы. Для каждого элемента, используемого для сопоставления, например, кодирование дат, должно быть согласования во всех наборах данных.

Все совпадающие данные должны иметь согласованные характеристики, размеры наборов данных и статус доступа (блокировки).

Чтобы компенсировать ошибки в наборах, сопоставление данных используются фонетические алгоритмы и компараторы строк. Фонетические алгоритмы справляются с орфографическими ошибками. Если слова будут содержать один и тот же фонетический код, поскольку они звучат почти одинаково и, следовательно, будут считаться согласованными, если их сравнивать по фонетическому коду, например, “напряжение – натяжение”, “ток – поток – моток”.

Для сопоставлений по фонетическому коду используются алгоритмы Soundex и Metaphone [6]. Однако Soundex разработан для иностранных слов, он менее подходит, так как игнорирует часть звуков русского языка. Использование Soundex может привести к тому, что данные будут согласовываться, хотя на самом деле они разные. Существуют алгоритмы методов сопоставлений по фонетическому коду, которые были разработаны для русского языка (Metaphone). Например, поиска контрагентов по фонетическому коду в базе данных (БД) программы 1с.

Следующим инструментом для применения может быть компаратор строк как инструмент для борьбы с ошибками, учитывающий отсутствующие или лишние элементы данных, и транспозиции (когда элемент ошибочно оказывается слева или справа) в наборах данных. Компаратор строк - метрика от 0 до 1, где 1 обозначает идеальное согласие, а 0 - совершенное несогласие. Джаро представил компаратор строк, который показал свою надежность при использовании для сопоставления данных. Алгоритм основан на длине двух наборов данных и количестве общих элементов. Винклер обнаружил, что в начале строки совершается меньше ошибок, чем в ее конце, и поэтому улучшил компаратор строк Джаро, введя веса. В целом, компаратор строк известен как Джаро-Винклера [7]. Метод применен в разработанном алгоритме.

Исследования показали, что при сравнении длинных строк, наилучший результат показывает компаратор строк Джаро-Винклера при сравнении короткой строки с длинной, и получении результата по основанию длинной строки. Преимущества компаратора строк Джаро-Винклера: работает с разной длиной строк; высокая точность сопоставления; нормированный результат (то есть от 0 до 1).

Сопоставления данных, согласно проведенным исследованиям выполняются на уровнях:

- коррекционном – процесс обеспечения сопоставимости данных, направлен на устранение неоднородности собранных данных и дальнейшее преобразование информации в согласованный, единый формат, позволяющий проводить прямые сравнения между различными наборами данных;
- экспертном - процесс обеспечения сопоставимости данных направлен на формирование оптимальной выборки изделий ЭКБ для применения в РЭА.

Указанные результаты исследований и требования по подготовке данных к сопоставлению применены и соблюдены в разрабатываемом алгоритме.

### Результаты исследований алгоритмов сопоставления данных

Алгоритмы сопоставления данных используются для определения существенные общих черт анализируемой информации. Они учитывают вариации, орфографические ошибки и другие несоответствия, чтобы выявить общие черты в данных для отнесения их к тем или иным группам ЭКБ.

Существует два основных подхода к автоматизации процессов сопоставлению данных:

а) точное сопоставление: данные сопоставляются, если элементы в них совпадают или почти совпадают;

б) неточное сопоставление: вычисляется оценка сходства между данными; данные сопоставляются, если их оценка сходства превышает указанный порог.

На основании перечисленных подходов определены три основные группы алгоритмов решающие задачи сопоставления данных:

1) точное сравнение (детерминированные алгоритмы) полагаются на определенные шаблоны и правила для назначения весов и оценок для определения сходства;

2) неточное сравнение (алгоритмы вероятностного сопоставления) полагаются на методы для оценки вероятности того, исследуемые наборы данных представляет одну и ту же сущность;

3) обучающие. Группа образовалась с развитием технологий искусственного интеллекта. Применяются алгоритмы машинного обучения (МО). Они используют широкий вектор признаков для расчета оценки сходства. С помощью оптимизации выполняется определение весов для этого расчета в соответствии с эталонным обучающим набором.

Среди вышеперечисленных групп исследовались алгоритмы сопоставления данных обеспечивающие ожидаемый результат. Установлено: слишком строгое правило соответствия дает пропуски действительных совпадений; свободные правила логически объединяют несвязанные данные.

Результаты исследований показали необходимость разработки алгоритма, комбинирующего методы: точные сопоставления, вероятностные сопоставления и сложные сопоставление, учитывающие существенные различия в данных и использующие теологии МО.

Детали реализации значительной части используемых алгоритмов и описания большинства методов не доступны. Это делает невозможным их полноценное использование для выполнения поставленной задачи. Поэтому разработка алгоритма сопоставления данных об ЭКБ в Системе является актуальной и в настоящее время выполнена. Алгоритм применен в Системе.

Алгоритм получился комбинирующим. Он состоит из частей: на основании МО; вероятностного и точного сопоставления. Результаты разработки частного алгоритма точного сопоставления представлены в этой статье.

### Разработка алгоритма точного сопоставления данных

Разработке предшествовало тщательное изучение постановки задачи, отечественного и зарубежного опыта в данной области, применения алгоритмов подобного назначения. Разработка алгоритма точного сопоставления данных включала ряд особенностей, выделенных в данном разделе.

**Концепция** разрабатываемого алгоритма точного сопоставления данных предусматривает сопоставление элементов наборов данных только в том случае, если они совпадают. Такое использование детерминированного подхода к сопоставлению значительно снижает вычислительную нагрузку и обеспечивает получение положительного результата.

Алгоритм точного сопоставления данных (СИД) сравнивает элементы данных, которые должны быть идентичными для сопоставления. Точное сопоставление идеально подходит для работы с четко определенными атрибутами, но оно может оказаться недостаточным при наличии вариаций, опечаток или неточных данных.

Данные сравниваются, и совпадения определяются. При положительном результате результат откладывается в сторону, и выполняется следующее сопоставление через остаточные наборы данных. Это выполняется до тех пор, пока сопоставляемые наборы данных не закончатся.

Понятия точности совпадения предполагают, что данные совпадают, а правила просты и понятны. Информация, поступающая от нескольких источников и в нескольких форматах может иметь пропуски или быть неточной, а также могут быть различия точности в данных. В результате приоритет от-

дается предсказуемости, а не точности. Это очень важно для операционных сценариев использования, таких как ассоциирование параметров изделий ЭКБ с требованиями к их применению, но в конечном итоге недостаточно для большинства других сценариев использования. Так, например, показатель наработки до отказа, утверждает, что изделие с высокой вероятности будет работать и выполнять заявленные изготовителем функции до указанной величины срока эксплуатации. Применение изделия ЭКБ в РЭА с меньшим сроком эксплуатации возможно, но не всегда целесообразно, если есть альтернативы.

Еще одна важная особенность алгоритма точного сопоставления заключается в том, что СИД обеспечивает соответствие один к одному, и есть только два возможных исхода решения: совпадение или отсутствие совпадения. Проведённые исследования показывают целесообразность проведения процедуры СИД перед переходом к вероятностному сопоставлению данных, если цель состоит в том, чтобы выполнить взаимно однозначное сопоставление наборов совпадающих данных, поскольку это может снизить общую вычислительную нагрузку.

Существует три возможных сценария использования СИД:

1) Если наборы данных сопоставляются по всем рассматриваемым элементам, маловероятно, что это случайно. Предполагается, что рассматриваемые наборы данных принадлежат одному и тому же изделию;

2) Если наборы данных не сопоставляются, маловероятно, что рассматриваемые наборы данных принадлежат одному и тому же изделию;

3) Если существуют промежуточные ситуации, когда некоторые элементы наборов данных сопоставляются, а некоторые не сопоставляются, необходимо доказать, принадлежат ли они одному и тому же изделию.

**Правила** для алгоритма точного сопоставления данных разрабатываются, апробируются и непрерывно тестируются. Это позволяет убедиться в их функционировании должным образом по мере поступления данных в Систему для выполнения задач сопоставления. Данные, с характеристиками изделий, отличные от первоначально получаемых, могут потребовать доработку набора правил. Требования к правилам обеспечивают выполнения задач сопоставления данных с минимальным числом ошибок, для исключения получения неверных и пропуска верных результатов.

Снижение качества данных или увеличение их сложности может привести к увеличению числа правил, необходимых для сопоставления данных. В конечном счете, эти правила могут стать многочисленными и не взаимосвязанными. Их трудно будет выполнить.

СИД эффективно используется, когда элементы данных согласованы (внутренне непротиворечивы) и структурированы. Наборы данных идентичные, если их значения совпадают. Методы СИД позволяют связывать элементы в наборах данных. При выполнении алгоритма СИД незначительные изменения в формате данных могут привести к несоответствию. Этому способствует неоднородность данных, которая обнаруживается на этапе их первичного анализа. И решается с помощью анализа распределения данных, который реализуется алгоритмом на основании расчетов (1):

$$S_{average} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (1)$$

где,  $x_i$  – значение элемента в наборе данных  $i$ -го источника,  $n$  – количество источников данных.

Весомый вклад в результаты работы алгоритма СИД внесён синтаксическим подходом. Методы синтаксического подхода определяют сходство между элементами, основываясь на эквивалентности наборов данных. Наиболее простой метод для определения сходства между элементами – точное совпадение наборов данных. Примером алгоритмов СИД являются алгоритм двоичного поиска, методы которого использованы в разработке.

В разработанном алгоритме применен набор правил каскадного типа: поиск по условному обозначению изделия. Если совпадения есть, поиск по комбинациям классификационных кодов изделия. Если совпадение есть, поиск по параметрам и показателям. Предполагаемое преимущество заключается в том, что правила предсказуемы.

Еще одним решением для совпадения является таблица оценок, которая взвешивает различные части данных об изделии и создает ряд правил. Это похоже на концепцию каскадных правил, но с

большим количеством возможностей для тонкой настройки.

Например, если имеется величина коэффициента усиления для микросхемы операционного усилителя, это не считается знанием того, какое у микросхемы условное обозначение. Несмотря на то, что конкретная величина коэффициента усиления для микросхемы операционного усилителя не совсем распространена на всю группу, есть по крайней мере несколько других микросхем с такой же величиной для коэффициента усиления. Таблица баллов позволяет затем присвоить баллы для каждого типа совпадающих данных и установить пороговое значение минимального объема информации, необходимого для уверенного сопоставления изделий. Например, совпадение условных обозначений - 3 балла, совпадение функционального назначения - 1 балл, совпадение кодов классификации - 4 балла, совпадение наименований параметров - 1 балла, совпадение величины параметров - 1 балл. Итого – точное совпадение – 9 баллов. Если установить порог – выше 6 баллов, это может дать результаты, которые показаны в Таблицы 1.

Таблица 1

### Результаты установления порога баллов

Баллы сопоставления	Обозначение	Назначение	Коды	Параметры	Величины
Обозначение	3	4	7	4	4
Назначение	4	1	5	2	2
Коды	7	5	4	5	5
Параметры	4	2	5	1	2
Величины	4	2	5	2	1

Только сопоставление всех вышеперечисленных элементов дает положительный результат. При этом, сопоставление обозначения и кода даст результат выше порога. Для сопоставления наименований параметров и величин параметров в группу, для сравнения, выше порога должны попасть коды классификации.

Правила сопоставления могут наращиваться в силу специфики для отдельных наборов данных изделий ЭКБ, развития технологий разработки, изготовления и применения.

**Идентификаторы** создаются для реализации алгоритма точного сопоставления данных и должны однозначно идентифицировать информацию об изделиях ЭКБ, для сопоставления от различных источников данных.

В качестве основного идентификатора, определено обозначение изделий ЭКБ. Оно присваивается изделию в общем порядке на основании ГОСТ Р 2.201-2023 “Единая система конструкторской документации. Обозначение изделий и конструкторских документов”, или для изделий с более ранним сроком изготовления по одноимённому ГОСТ 2.201-80. Для отдельных классов изделий существуют особые правила присвоения обозначений. Например, система обозначений интегральных микросхем установлена ОСТ 11 073.915-2000.

Обозначения изделия является неизменным в течение всего жизненного цикла и состоят из букв кириллицы и цифр. В отдельных обозначениях могут присутствовать знаки пунктуации. В редких случаях в обозначениях используется латиница. Однако, зарубежные аналоги для отечественных изделий ЭКБ указываются латиницей. Автоматизация процессов подбора, которых исследована автором и используется в настоящей работе [8].

Обозначения могут дублироваться. Изделия могут иметь одинаковые обозначения, но принадлежать разным классам изделий. Поэтому, в качестве вспомогательного используется классификатор изделий и конструкторских документов ОК 012.

В процессе разработки определено, что данные, имеющие отношение к выбранному идентификатору, относятся к данному изделию, в то же время данные, не имеющие отношение к выбранному идентификатору, возможно идентифицируют другие изделия.

**Блокирующие переменные** созданы и применяются в алгоритме для определения направления сопоставления данных.

Результаты исследований показали необходимость включения блокирующей переменной, определяющего направление сопоставления данных в зависимости от типа информации. Информация об изделиях ЭКБ, получаемая от источников в процессах сбора, делится на следующие типы:

- первичная;
- дополняющая;
- корректирующая.

Сбор первичной информации выполняется по вновь разрабатываемым изделиям ЭКБ, или изделиям, информация о которых отсутствовала в Системе. Основные требования к сбору первичной информации: она собирается в строгом соответствии с целями задачи сбора, методология сбора информации строго контролируется.

К первичной информации относят информацию, отсутствие которой влечет за собой потерю дальнейших сведений об изделии, описаний и нарушение логики восприятия информации, в которой выражена коммуникативная задача представления информации об изделии.

Первичная информация размещается в начальных наборах данных об изделии ЭКБ. В состав первичной информации входят:

- условное обозначение изделия;
- наименование разработчика изделия;
- наименование изготовителя изделия;
- наименование калькодержателя документации;
- классификационные коды изделий;
- функциональное назначение;
- область применения изделия;
- текущая стадия жизненного цикла изделия;
- параметры и показатели изделия, другие требования к изделию, необходимые для принятия

решения на ее применение. Они должны соответствовать требованиям документов по стандартизации на конкретные группы изделий.

Дополняющая информация выполняет целый ряд существенных функций, являясь объектами адекватного понимания информации об изделии, так как она конкретизирует, уточняет первичную информацию, раскрывает ее содержание, дополняет ее примерами, цифрами, фактами, графиками, диаграммами, иллюстрациями, ссылками на другую информацию, необходимую для детального понимания особенностей функционирования изделий.

К дополняющей относится конкретизирующая, резюмирующая, и иллюстрирующая информация. Дополняющая информация расширяет и углубляет первичную информацию путем ее разъяснения, толкования, раскрытия сущности изделия. В дополняющей информации приводятся конкретные примеры, уточняющие, подтверждающие, или объясняющие свойства ЭКБ.

Корректирующая информация позволяет привести уточнения первичной и дополняющая информации об изделии.

На практике, дополняющая и корректирующая информация связаны с друг-другом, решая конкретный круг задач – увеличение полноты информации, необходимой и достаточной для принятия решения на выбор изделия.

Информация собираются в БД, практическая реализация сбора данных выполнена автором и подтверждена в [9,10]. Практическая реализация процедуры контроля данных на уровне визуализации подтверждена в [11].

Описанные выше типы информации применены в качестве блокирующих переменных: первичная; дополняющая; корректирующую.

Определение того, где следует устанавливать блокирующие переменные, представляет собой балансировку между получением приемлемой чувствительности (доли действительно совпадающих данных, связанных алгоритмом) и положительной прогностической ценностью (или точностью, долей данных, связанных алгоритмом, которые действительно совпадают). Для получения наилучших значений, в данных исследованиях применены блокирующие переменные. Они позволяют ограничить срав-

нения только теми наборами данных, для которых согласуются идентификаторы, что приводит к увеличению положительной прогностической ценности в ущерб чувствительности.

Блокирующие переменные должны быть небольшого размера, чтобы избежать слишком большого количества непродуктивных сравнений, но достаточными, чтобы предотвратить распространение данных для одного и того же изделий и таким образом невозможность получения верных совпадений.

После того, соответствующие элементы данных определены для текущей итерации и блокирующей переменной, их наборы данных должны быть определены и запротоколированы как сопоставленные.

Значения блокирующей переменной типов информации  $TYPE_{information}$  представлено в определении (2):

$$TYPE_{information} = \begin{cases} 1, & \text{если тип информации – первичная} \\ 2, & \text{если тип информации – дополняющая} \\ 3, & \text{если тип информации – корректирующая} \end{cases} \quad (2)$$

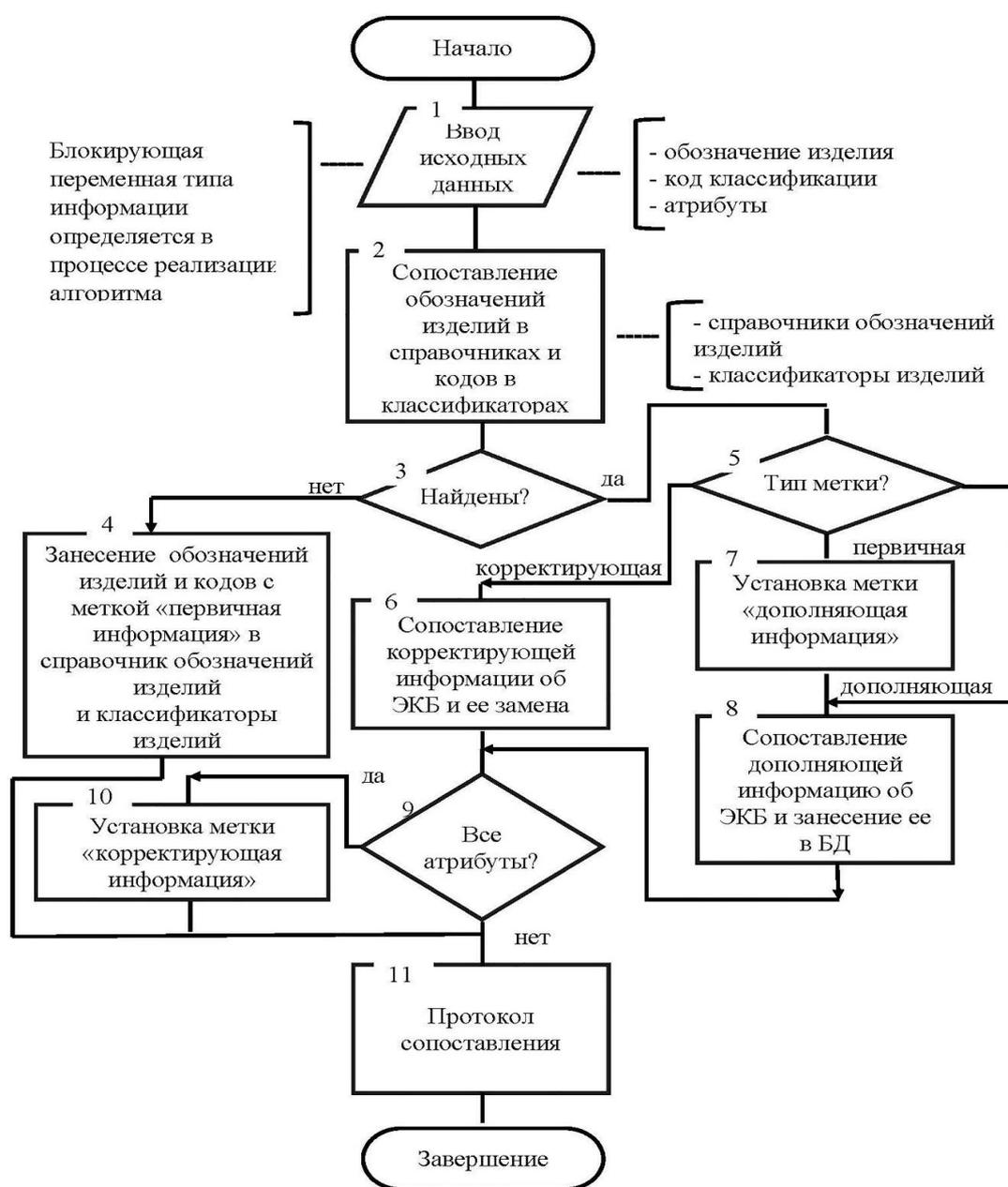
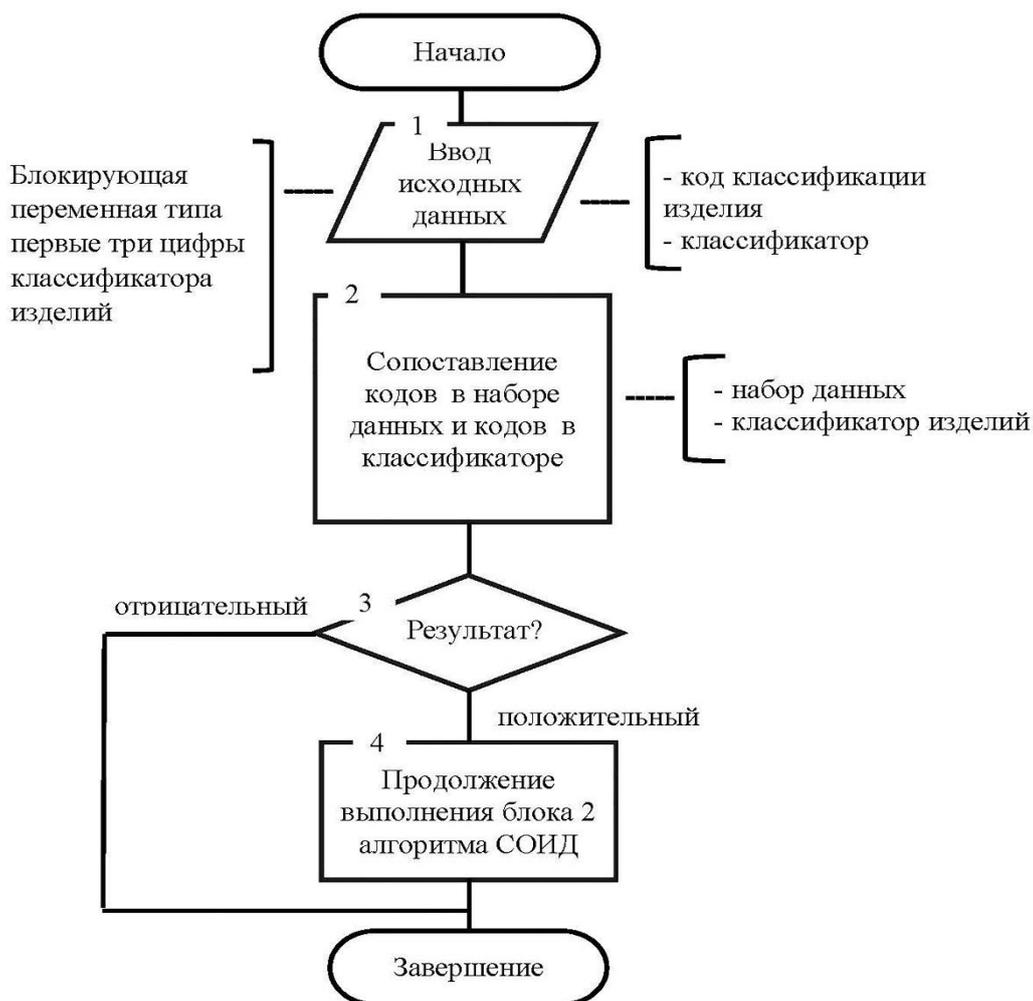


Рис. 4. Частный алгоритм определения идентификаторов точного сопоставления с применением блокирующей переменной типа информации



**Рис. 5. Алгоритм реализации применения классификационного кода в виде блокирующей переменной**

Блокирующая переменная типа информации, направленная на сокращение пространства поиска между наборами данных, избегая сравнения пар элементов, которые с наименьшей вероятностью будут совпадениями. Например, если оба набора данных содержат 10 элементов и необходимо провести сопоставление один к одному, то получится 100 элементов для сравнения. Это без блокирующих переменных значительно увеличивает требуемые ресурсы.

Частный алгоритм определения идентификаторов точного сопоставления с применением блокирующей переменной типа информации представлен на рис. 4.

В представленном на рис. 4 алгоритме, пространство совпадения сокращено путем сопоставления только тех данных, которые потенциально могут быть совпадениями, за счет блокирующей переменной типа информации.

Для итеративных проходов по наборам данных используются несколько блокирующих переменных. В алгоритме, представленном на рис. 4 блокируются сопоставление по типу информации, а также по классификационному коду. Первые цифры классификационного кода сокращают пространства совпадений. Используемый в алгоритме классификатор изделий и конструкторских документов ОК 012 – Классификатор ЕСКД, представляет собой систематизированный свод наименований классификационных группировок объектов классификации – в том числе изделий ЭКБ и является составной частью системы классификации и кодирования технико-экономической информации. В Классификаторе ЕСКД включены классификационные характеристики изделий, на которые разработана и разрабатывается конструкторская документация по ЕСКД.

Структура обозначения кода классификационной характеристики имеет вид: 99 9 9 9 9. Первая и вторая цифры – класс, третья цифра – подкласс, четвертая цифра – группа, пятая цифра – подгруппа, шестая цифра – вид.

В классификаторе, 43 класс – микросхемы, приборы полупроводниковые, электровакуумные, пьезоэлектрические, квантовой электроники, резисторы соединители, преобразователи электроэнергии, средства вторичного электропитания. Это не вся номенклатура ЭКБ. Реле, например, находятся в другом классе.

Если вернуться к 43 классу, то можно сузить объём сопоставления данных за счет подкласса, например - 431 микросхемы интегральные.

Реализация применения классификационного кода в виде блокирующей переменной представлена в алгоритме на рис. 5. Алгоритм является частью алгоритма определения идентификаторов точного сопоставления, представленного на рис. 4, блок 2.

В алгоритме данные сопоставляются и результаты откладываются, а затем сопоставляются оставшиеся наборы данных, используя другие критерии блокировки, такие как параметры и показатели. Учитывая их сложность, используются другие вышеописанные алгоритмы: вероятностного сопоставления и с применением МО.

**Частота обновления данных** исследована и используется в разработанном алгоритме СИД. Сопоставляемые элементы наборов данных могут иметь различную частоту. Данными с высокой частотой обновления являются наборы с большим количеством значений, например, параметры изделий ЭКБ при различных условиях эксплуатации. Примерами наборов данных с небольшой частотой изменений являются данные о наработке до отказа изделий ЭКБ. Данными с нулевой частотой обновления являются обозначения изделия, классификационные коды и функциональное назначения. В алгоритмах сопоставление необходимо учитывать такую особенность частоты обновления данных. Применение такого показателя существенно сокращает затраты на сопоставление.

Соглашения к сопоставлению элементов набора данных по частоте обновления следующие:

- обозначение, классификационные коды изделия ЭКБ, функциональное назначение не меняются в течение жизненного цикла изделия;
- область применения, определяется при создании, очень редко дополняется;
- параметры и показатели изделия, конструктивное исполнение могут меняться в течение жизненного цикла изделия.

Таким образом, перед применением алгоритмов вероятностного сопоставления и с применением технологий МО, целесообразно провести процедуру СИД, чтобы выполнить взаимно однозначное сопоставление набора совпадающих переменных, поскольку это может снизить общую вычислительную нагрузку Системы.

### Параметры алгоритма точного сопоставления данных

Параметрами алгоритма точного сопоставления данных являются: точность, эффективность, масштабируемость, гибкость и интеграция.

1) точность, определяет количество правильных совпадений из общего числа обработанных данных. Точность показывает близость вариантов, сопоставляемых данных к истинному результату.

2) эффективность, определяет объёмы ресурсов необходимые для выполнения задач СИД. Основными ресурсами эффективности алгоритма СИД считаются: время выполнения алгоритма и объём используемой памяти. Эффективность зависит от числа и сложности выполняемых алгоритмом операций, вычислительных ресурсов, объёмов сопоставляемых данных; количества переменных, типа и размеров структуры данных, вызовов функций и способа выделения памяти. Для повышения требуемой эффективности необходимо тестировать алгоритм СИД при увеличении входных данных, оптимизировать программный код и разрабатывать оптимальные структуры входных данных.

3) масштабируемость - способность решать задачи СИД с увеличением рабочей нагрузки при добавлении ресурсов без структурных изменений в Системе. Под добавлением ресурсов понимается создание параллельных процессов алгоритмом. Масштабируемость показывает способность алгоритма выполнять сопоставление больших объёмов данных об ЭКБ, учитывая их потенциальный рост. В алгоритмах с плохой масштабируемостью добавление ресурсов приводит лишь к незначительному повы-

шению производительности.

4) гибкость – предполагает настраиваемые правила сопоставления, когда алгоритм работает с различными форматами и структурами данных, это: одномерные, многомерные, динамические и ассоциативные массивы; связанные списки; стеки; очереди; множества; префиксные деревья и деревья поиска; ориентированные и неориентированные графы.

5) интеграция - определяет совместимость алгоритма СИД с существующими рабочими процессами и программными средствами Системы: обработку информации от нескольких источников с различными форматами; очистки и стандартизации данных; создания словарей в Системе; установки атрибутов и выполнение требований к: релевантности для выявления дубликатов или сходств; варьированности для неточного совпадения и оценки совпадений.

Параметры СИД могут изменяться от минимальной величины близкой к нулю, до единицы. При умножении на 100%, параметры СИД могут показывать значение в процентах.

### Заключение

В настоящей работе исследована применимость СИД в Системе. На основании результатов исследований определены требования к данным. Разработан алгоритм СИД, основанный на принципах определенности. Он полагается на идентификаторы и атрибуты, чтобы связать данные об ЭКБ из различных источников. Используя уникальные идентификаторы, алгоритм не оставляет места для двусмысленности, предлагая точный и надежный метод сопоставления данных. Следовательно, СИД хорошо справляется с обеспечением точной агрегации данных на конкретном уровне. Это позволяет с высокой степенью уверенности создавать комплексные наборы данных об изделиях ЭКБ. Однако алгоритм СИД сталкивается с ограничениями в сценариях, где отсутствуют конкретные идентификаторы, что приводит к рискам разрозненности данных и неполным наборам данных об изделиях ЭКБ.

Самое очевидное преимущество алгоритма СИД перед его аналогами заключается в том, что он обеспечивает точное и надежное сопоставление при наличии точных и надежных элементов данных. СИД не создаст совпадение, если элемента нет, и в результате определит в среднем меньше совпадений, чем вероятностные, или МО. Но, для случаев, когда точность данных является наивысшим приоритетом, алгоритм СИД будет предпочтительнее.

Было неправильным утверждать, что один из типов алгоритмов СИД, вероятностный, или МО превосходит другой, поскольку каждый из них имеет свои сильные и слабые стороны в определенных обстоятельствах. Система использует комбинацию как СИД, так и вероятностного алгоритмов сопоставления, методов с применением МО, поэтому выбор между ними является в некоторой степени ложной дихотомией. СИД применяется к элементам с высокой степенью достоверности, в то время как вероятностное и с МО сопоставление используется в случаях, когда идентификаторы являются неполными, несогласованными или отсутствующими. Используя гибридный подход, Система максимизирует точность и гибкость, создавая более полное сопоставление данных об ЭКБ.

Исследования показывают, что уровень оптимального формирования предложений на применение изделий ЭКБ выходит за рамки банального анализа относительных показателей. Он плавно переходит к оценке статистической значимости параметров изделий ЭКБ, где вместо систематической проверки неоднородности данных, оценки их качества и применимости, происходит нахождение сходств или отличий между элементами наборов данных, составляющими единое целое, а также изучение факторов, которые могут влиять на результаты выборки изделий. Например, данные изделий могут быть неоднородны по уровню стойкости к внешним воздействиям, группы изделий могут различаться по относительной факторной применимости, а классы изделий по функциональному назначению. Такой уровень сопоставимости должен будет указывать на эти различия внутри отдельных выборок, между ними и между результатами в анализе свойств изделий.

Проблемы при сопоставлении данных возникает в тех случаях, когда возникают ошибки в сопоставлении элементов и уникальный идентификатор отсутствует. В этом случае используется вероятностная увязка данных с соответствующими элементами, которые могут быть подвержены ошибкам.

Исключения, такие как отсутствующие идентификаторы, включает в себя создание дополнитель-

ных правил сопоставления данных. Одним из таких правил в случае отсутствия идентификатора может быть сравнение классификационных кодов в целях найти совпадение.

В исследованиях методов и разработке алгоритмов вероятностного, и МО такие выводы использованы и учтены.

#### Список источников

1. Дормидошина Д.А., Рубцов Ю.В., Применение моделей электронной компонентной базы. Радиоэлектронная отрасль: проблемы и их решения. 2024. № 14 - URL: <https://www.deyton.ru/doc/stat22.07.pdf> (дата обращения: 25.01.2026).
2. Алексеева А.В., Рубцов Ю.В., Разработка IBIS-модели микросхем концентраторов сетей, применяемых для проектирования телекоммуникационной аппаратуры. "Нано- и микросистемная техника" Т. 19. № 6. г. Москва, 2017 - URL: [https://www.deyton.ru/doc/ibis\\_10.07.2017.pdf](https://www.deyton.ru/doc/ibis_10.07.2017.pdf) (дата обращения: 25.01.2026).
3. Дормидошина Д.А., Савин М.Л., Рубцов Ю.В., Применение ИСМН в процессах сбора, обработки и анализа информации о надежности изделий микроэлектроники. "Нано- и микросистемная техника", Том 22, №9, г. Москва 2020 - URL: <https://www.deyton.ru/doc/ISSN18138586.pdf> (дата обращения: 25.01.2026).
4. Гагарина Л. Г., Рубцов Ю.В., Особенности разработки метода классификации плоских QFN-корпусов для применения в составе автоматизированных систем технической подготовки производства изделий микроэлектроники. "Известия высших учебных заведений. Электроника". 27(3). Москва, 2022 - URL: <https://www.deyton.ru/doc/QFNRubtsov.pdf> (дата обращения: 25.01.2026).
5. Дормидошина Д.А., Савин М.Л., Рубцов Ю.В., Исследования и результаты сбора, анализа и обобщения информации о выводных рамках, применяемых в изделиях микроэлектроники. "Нано- и микросистемная техника", Том 22, №7, г. Москва, 2020 г. - URL: <https://www.deyton.ru/doc/CRENNAS.pdf> (дата обращения: 25.01.2026).
6. В. С.Вывоханец, Ду Цзяньмин, Сакулин С.А. Обзор алгоритмов фонетического кодирования. 2018. (дата обращения: 06.05.2025). <https://cyberleninka.ru/article/n/obzor-algoritmov-foneticheskogo-kodirovaniya> (дата обращения: 25.01.2026).
7. Воронина И.Е., Экерт Н.А. Выбор варианта из множества решений при нечетком сравнении строк. Воронежский государственный университет. Компьютерная лингвистика и обработка естественного языка. 2023 - URL: <https://journals.vsu.ru/sait/article/view/11428> (дата обращения: 25.01.2026).
8. Рубцов Ю.В., Оптимизация процессов подбора аналогов изделий электронной компонентной базы. Электронная техника. Серия 3: Микроэлектроника. 2015- URL: [https://www.niime.ru/upload/zhurnal-mikroelektronika/%D0%92%D1%8B%D0%BF%D1%83%D1%81%D0%BA%203%20\(159\)%202015.pdf](https://www.niime.ru/upload/zhurnal-mikroelektronika/%D0%92%D1%8B%D0%BF%D1%83%D1%81%D0%BA%203%20(159)%202015.pdf) (дата обращения: 25.01.2026).
9. Свидетельство о государственной регистрации базы данных №2015621293 Российская Федерация, Корпуса ЭКБ: заявка №2015620803, дата поступления 26.06.2015: дата государственной регистрации в Реестре баз данных 26.08.2015/ Грязнова Т.В., Довгань И.Д., Рубцов Ю.В. заявитель АО "ЦКБ "Дейтон".
10. Свидетельство о государственной регистрации программы для ЭВМ №2022683437 Российская Федерация, Программа формирования базы данных об изделиях электронной техники: заявка №2022682511, дата поступления 22.11.2022: дата государственной регистрации в Реестре программ для ЭВМ 05.12.2022/ Рубцов Ю.В., Дормидошина Д.А., Криницкий В.В., Шишкова Ю.М., Окунев К.Е.; заявитель АО "ЦКБ "Дейтон".
11. Свидетельство о государственной регистрации программы для ЭВМ №2023614155 Российская Федерация, Программа визуализации данных об изделиях электронной техники: заявка №2023612746, дата поступления 14.02.2023: дата государственной регистрации в Реестре программ для ЭВМ 27.02.2023/ Рубцов Ю.В., Дормидошина Д.А., Криницкий В.В., Курилов А.В., Окунев К.Е.; заявитель АО "ЦКБ "Дейтон".

**НАУЧНОЕ ИЗДАНИЕ**

**НАУЧНОЕ ОБОЗРЕНИЕ: АКТУАЛЬНЫЕ ВОПРОСЫ  
ТЕОРИИ И ПРАКТИКИ**

Сборник статей

Международной научно-практической конференции

г. Пенза, 10 февраля 2026 г.

Под общей редакцией

кандидата экономических наук Г.Ю. Гуляева

Подписано в печать 11.02.2026.

Формат 60×84 1/16. Усл. печ. л. 18,9

МЦНС «Наука и Просвещение»

440062, г. Пенза, Проспект Строителей д. 88, оф. 10

[www.naukaip.ru](http://www.naukaip.ru)

