



# ЦИФРОВАЯ КАТАЛОГИЗАЦИЯ: СТАНДАРТНЫЕ И НОВЫЕ МЕТОДОЛОГИИ

## DIGITAL CATALOGING: STANDARD AND NOVEL METHODOLOGIES

УДК 025.3

The author reviews product cataloguing methodologies, highlights their advantages and disadvantages, and presents a new approach using artificial intelligence technologies that opens up broad prospects for market participants.



**В последние годы активно растет интерес к цифровой каталогизации продукции и услуг со стороны участников рынка — бизнеса и государства. Все большее применение находят и стандартные методологии (которые имеют как преимущества, так и ограничения), и новые подходы с использованием технологий искусственного интеллекта.**

Цель каталогизации — обеспечение единства описания и идентификации продукции и услуг для поддержки бизнес-процессов по снабжению (закупки, логистика, управление складскими запасами) между участниками рынка.

В настоящее время существует ряд основополагающих проблем, которые требуют незамедлительного решения. В частности, отсутствуют централизованный информационный ресурс по каталогизации продукции и услуг с единым идентификатором, общая методология, стандартизированные наборы характеристик продукции, в силу чего поиск ее аналога сильно затруднен.

Также следует отметить, что повсеместно используются стандартные методологии, подразумевающие ручной способ обработки и формирования каталогов, что значительно увеличивает финансовую нагрузку на

бюджет участников рынка за счет затрат на фонд оплаты труда и не сокращает сроки работ по каталогизации. Применяются частично или не применяются вовсе технологии искусственного интеллекта (ИИ) для ускорения и точности обработки данных и создания каталогов, т.к. нет стандартизированных датасетов и баз знаний.

Существующие методологии каталогизации имеют свою специфику, ограничения и преимущества для бизнеса и экономики в целом.

### НАЦИОНАЛЬНАЯ КАТАЛОГИЗАЦИЯ

На текущий момент в силу отсутствия централизованного национального ресурса на рынке представлены отдельные государственные системы, которые направлены либо на реализацию государственной политики, либо на обеспечение государственных нужд.



**Ключевые слова:** каталогизация, методологии, цифровая каталогизация, машинное обучение, искусственный интеллект.  
**Keywords:** cataloging, methodologies, digital cataloging, machine learning, artificial intelligence.

Рассмотрим кратко действующие информационные ресурсы по каталогизации продукции.

## Каталог продукции Государственной информационной системы промышленности (ГИСП)

*Назначение:* ведение каталога промышленной продукции, произведенной на территории Российской Федерации в соответствии с постановлением Правительства РФ от 17 июля 2015 г. № 719 «О подтверждении производства промышленной продукции на территории Российской Федерации».

*Идентификатор:* уникальный цифровой код.

*Базовый классификатор:* Общероссийский классификатор продукции по видам экономической деятельности (ОКПД2), Товарная номенклатура внешнеэкономической деятельности (ТН ВЭД).

## Цифровой паспорт промышленной продукции

*Назначение:* ведение цифровых паспортов для каталога промышленной продукции ГИСП согласно постановлению Правительства РФ от 29 июня 2024 г. № 896 «Об утверждении Правил формирования и использования цифровых паспортов промышленной продукции».

*Идентификатор:* уникальный цифровой код (дублируется в Каталог ГИСП).

*Базовый классификатор:* ОКПД2, ТН ВЭД.

## Каталог товаров, работ, услуг (ТРУ)

*Назначение:* формирование и ведение каталога товаров, работ, услуг для обеспечения государственных и муниципальных нужд в рамках Федерального закона от 5 апреля 2013 г. № 44-ФЗ «О контрактной системе в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд».

*Идентификатор:* уникальный цифровой код на основе кода ОКПД2.

*Базовый классификатор:* ОКПД2.

## Банк данных «Продукция России»

*Назначение:* формирование каталожных листов в рамках Реестра технических условий по Федеральному закону от 29 июня 2015 г. № 162-ФЗ «О стандартизации в Российской Федерации».

*Идентификатор:* регистрационный номер реестра.

*Базовый классификатор:* ОКПД2.

## Национальный каталог

*Назначение:* формирование каталога для обеспечения функционирования государственной системы цифровой маркировки товаров «Честный знак» в соответствии с постановлением Правительства РФ от 26 апреля 2019 г. № 515 «О системе маркировки товаров средствами идентификации и прослеживаемости движения товаров».

*Идентификатор:* глобальный номер товарной продукции (Global Trade Item Number, GTIN).

*Базовый классификатор:* ТН ВЭД.

## Федеральная система каталогизации продукции

*Назначение:* обеспечение единства описания и идентификации предметов снабжения согласно Федеральному закону от 29 декабря 2012 г. 275-ФЗ «О государственном оборонном заказе».

*Идентификатор:* федеральный номенклатурный номер.

*Базовый классификатор:* Единый кодификатор предметов снабжения для федеральных государственных нужд (ЕКПС).

## Проблемы и ограничения

Наряду с описанными выше ограничения в сфере национальной каталогизации заключаются в следующем. Отсутствуют открытые средства интеграции (API<sup>1</sup>) или открытые наборы данных, взаимодействие осуществляется только между государственными системами, причем правила ведения каталогизации определяются нормативными правовыми актами. Наборы характеристик применяются только к позициям, а не к классам продукции. В силу этого отсутствует наследование характеристик по классам — характеристики создаются многократно, и поиск аналога невозможен или содержит ошибки.

Применяемые базовые классификаторы также имеют ряд системных проблем. Объекты классификации сформированы без применения единого классификационного признака, по смешанным признакам (функциональному назначению, конструкции, характеристикам, марке и т.д.). Наименования объектов классификации только частично соответствуют документам по стандартизации (ДС). Классификационные отношения между объектами классификации построены в большей степени по ассоциации, а не по родовым и партитивным отношениям. Процедуры внесения изменений и создания новых объектов классификации занимают длительное время, проходят целый цикл согласований и публикаций и оформляются организационно-распорядительными документами.

## КОРПОРАТИВНАЯ КАТАЛОГИЗАЦИЯ

Поскольку отсутствует национальный информационный ресурс, а также нет единого подхода к каталогизации собственной продукции среди производителей, вертикально интегрированные компании вынуждены организовывать процессы и разрабатывать собственную методологию каталогизации.

Для устранения разночтений в терминологии термины «продукция», «товар», «услуга» и «материально-технический ресурс» далее в тексте статьи заменены на аббревиатуру НСИ (нормативно-справочная информация).

## Стандартная методология

Стандартная методология включает:

- организацию централизованного ведения НСИ;
- создание централизованной службы НСИ;

<sup>1</sup> API (Application Programming Interface) — интерфейс программирования приложений.

- применение корпоративных классификаторов;
- стандартизацию описания НСИ.

Ограничения стандартной методологии:

- обработка заявок на создание и изменение НСИ может занимать длительное время;
- сложная структура и взаимозависимость данных в корпоративных системах не позволяют оперативно менять структуру НСИ;
- атрибутивное описание (характеристики) НСИ не привязано к классам классификатора и не наследуется при создании НСИ в классе, характеристики дублируются и не управляются централизованно, поиск аналога по характеристикам невозможен или содержит ошибки;
- отсутствуют правила формирования стандартизованных наименований (шаблонов) НСИ из атрибутивного состава;
- возможно создание дублей НСИ при наличии синонимов и отсутствии словаря (гlossария) терминов;
- описание НСИ с помощью характеристик и классы не соответствуют ДС или не актуализируются при их обновлении;
- исходная НСИ может содержать характеристики и единицы измерений в другой системе стандартизации (ISO, EN, API, ASME, ASTM и т.д.), и при отсутствии конвертеров на национальную систему стандартизации

(ГОСТ/ГОСТ Р) могут создаваться дубли, поиск аналога по характеристикам невозможен или содержит ошибки.

### Расширенная методология

Преимущества расширенной методологии с применением машинного обучения и открытых языковых моделей ИИ:

- автоматическая и полуавтоматическая классификация НСИ на основе машинного обучения;
- автоматический и полуавтоматический поиск дублей НСИ и отклонение заявок на создание НСИ на основе машинного обучения;
- парсинг наименования НСИ и раскладка значений характеристик по атрибутам на основе машинного обучения;
- автоматическая и полуавтоматическая нормализация НСИ на основе машинного обучения;
- применение чат-ботов на основе открытых языковых моделей ИИ для формирования описания НСИ или создания уровней классификатора.

Ограничения расширенной методологии:

- обучение проводится на корпоративной модели данных (НСИ), которая может содержать ошибки, и, как следствие, алгоритмы будут обрабатывать также с ошибками (чем хуже качество данных, тем хуже будут результаты автоматической классификации и поиска дублей);

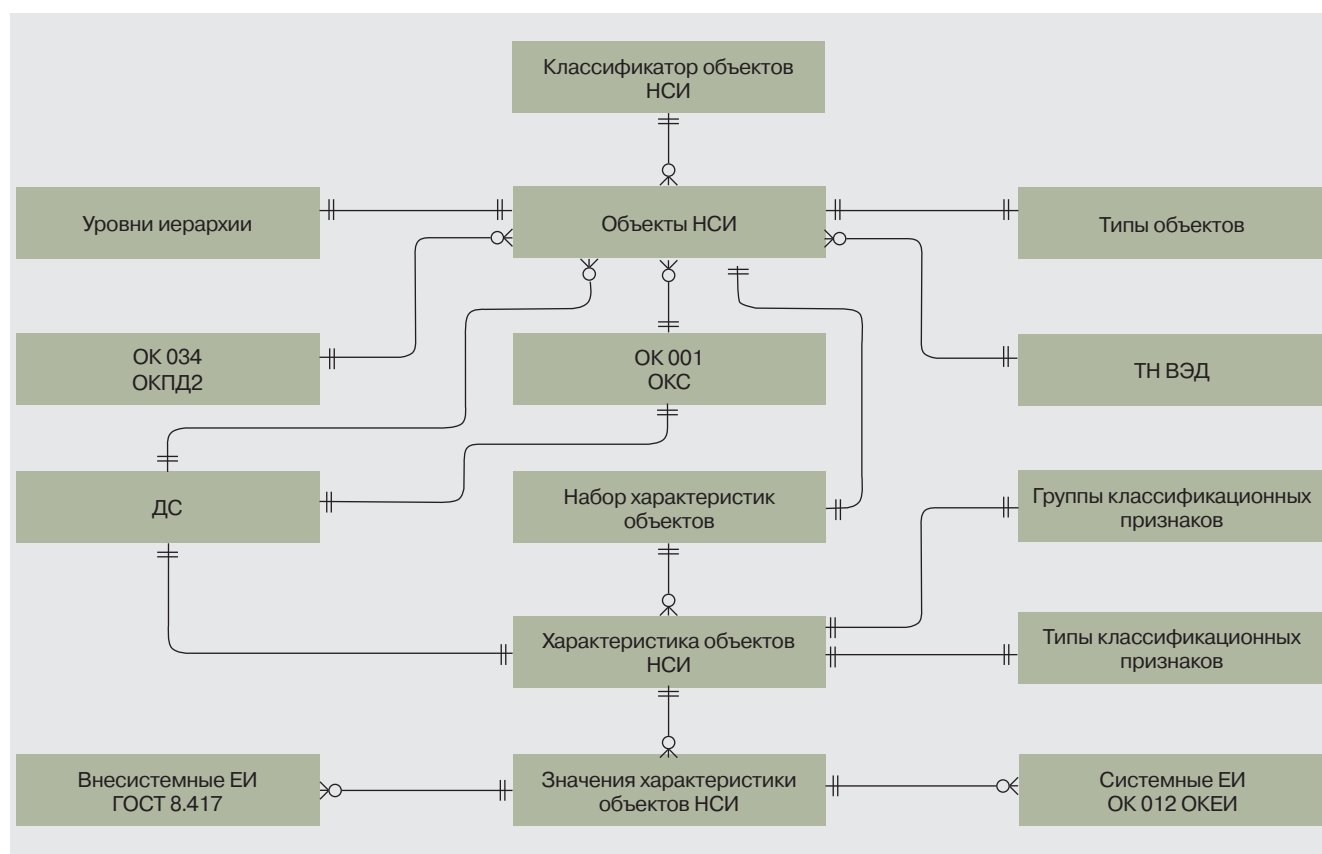


Рис. 1. Объектно-ориентированная модель данных.

НСИ — нормативно-справочная информация; ОК — объект классификации; ОКПД2 — Общероссийский классификатор продукции по видам экономической деятельности; ОКС — Общероссийский классификатор стандартов; ТН ВЭД — Товарная номенклатура внешнеэкономической деятельности; ДС — документы по стандартизации; ЕИ — единицы измерения; ОКЕИ — Общероссийский классификатор единиц измерения. ГОСТ 8.417—2024 «Государственная система обеспечения единства измерений. Единицы величин»



Рис. 2. Функциональная архитектура RSClass

- при отсутствии наборов данных по терминам и их синонимам, конвертеров характеристик и единиц измерений по различным системам стандартизации парсинг наименований и автоматическая нормализация НСИ будет отрабатываться с низкой точностью;

- чат-боты на основе открытых языковых моделей ИИ используют открытые и общие данные, доступные в интернете, но не имеют прямого доступа к конкретным документам или базам данных (доступа к актуальным и легитимным ДС у чат-ботов нет);

- ДС представляют собой неструктурированные данные, которые должны быть извлечены и структурированы (описание и классификация НСИ, сформированные чат-ботом, содержат ошибки, неточности и часто устаревшую информацию).

## НОВАЯ МЕТОДОЛОГИЯ И ПОДХОДЫ

Компания АО «ИндигоСофт ЦТ» апробировала на практике множество методологий по каталогизации, учла их ограничения и разработала собственную систему, стандартизовав ее на соответствие серии стандартов организаций. Сейчас она проходит научно-техническую экспертизу в ряде технических комитетов Росстандарта.

Методология базируется на применении объектно-ориентированной модели данных, которая состоит из объектного иерархического классификатора и фасетных

наборов данных (классификационные признаки). Общая структура приведена на рис. 1.

Для объектно-ориентированной модели данных применяется перечислительная система классификации с иерархической структурой и плавающим числовым методом кодирования. При построении иерархической структуры используется нисходящий метод, при котором иерархические уровни наследуют классификационные признаки и иные характеристики классов. Такая структура классификации формируется из композиции родовых и партитивных отношений.

Для обеспечения гармонизации объектного иерархического классификатора с общероссийскими и ведомственными классификаторами указывается связь (мэппинг) с их кодами. Классификатор и наборы классификационных характеристик НСИ формируются только из ДС и поддерживаются при актуализации и введении новых документов. Данная методология позволяет обеспечить соответствие классификации из ДС, применение стандартизированной терминологии и набора классификационных характеристик НСИ, гарантировать его наследование от верхнего до конечного уровня классификатора.

С целью моделирования данных и создания эталонных моделей НСИ для дальнейшей ее каталогизации по вышеуказанной методологии разработана система RSClass, схема функциональной архитектуры которой представлена на рис. 2.



Рис. 3. Функциональная архитектура ЕЦ НСИ

В настоящий момент апробируются алгоритмы машинного обучения по извлечению неструктурированной информации из ДС, поиску и подбору ДС по классификационным признакам, полуавтоматической классификации и формированию наборов характеристик НСИ.

Созданные эталонные модели данных могут применяться в качестве шаблонов описания НСИ, а также для машинного обучения и обучения языковых моделей ИИ в виде датасета. Для целей каталогизации НСИ на основе эталонных моделей, сформированных в RSClass, разработана система единой цифровой нормативно-справочной информации (ЕЦ НСИ), схема функциональной архитектуры которой приведена на рис. 3.

ЕЦ НСИ позволяет каталогизировать:

- эталонную позицию НСИ (единый идентификатор), которая содержит обезличенное наименование (без идентификационных элементов производителя), набор характеристик, коды общероссийских и ведомственных классификаторов и ее единицу измерения;
- позицию поставщика НСИ, которая содержит связь с эталонной позицией, идентификационные элементы производителя, ДС на поставку, цену и ее единицу измерения.

Функционал ЕЦ НСИ также предоставляет возможность:

- формировать шаблоны наименований эталонных позиций;
- автоматически подбирать позиции поставщиков к эталонной позиции;
- конвертировать характеристики и их единицы измерений из зарубежной в национальную систему стандартизации;
- интегрироваться с API и получать готовые наборы данных в стороннюю систему.

## ЗАКЛЮЧЕНИЕ

Методология и технологии АО «ИндигоСофт ЦТ» позволяют закрыть многие проблемы и потребности, а именно:

- обучить языковые модели ИИ на основе стандартизированных данных из ДС;
- настроить более точные алгоритмы машинного обучения на основе стандартизированных данных из ДС для выполнения операций с НСИ и снизить объем ручных операций;
- формировать каталоги НСИ на основе эталонных моделей данных;
- использовать указанные системы как источник данных для обогащения НСИ и повышения качества данных.

Рассмотренная методология и новые технологии открывают перед участниками рынка широкие перспективы. На их основе могут быть сформированы общая система национальной и межгосударственной каталогизации, единый межгосударственный каталог. Станет возможным автоматизированный обмен данными и обогащение НСИ национального и корпоративных каталогов для повышения качества формирования программ и планов снабжения, точного подбора аналогов. В перспективе планируется создание национальной специализированной языковой модели ИИ. Однако для успеха этой работы необходимо тесное взаимодействие и сотрудничество всех заинтересованных сторон.



**Виталий Викторович ЩУКИН** — генеральный директор АО «ИндигоСофт ЦТ»

**Vitaliy V. SHCHUKIN** — Director General, IndigoSoft CT JSC