

# Классификация данных с использованием средств визуализации

А.Е. Дзенгелевский<sup>1</sup>

Национальный исследовательский ядерный университет «МИФИ», Москва, Россия

<sup>1</sup> ORCID: 0009-0002-4770-1421, [Dzengelewski@gmail.com](mailto:Dzengelewski@gmail.com)

## Аннотация

В данной статье рассмотрены способы использования средств визуализации для построения классификаторов объектов в ходе автоматизации крупного предприятия. Предложенные подходы позволяют заинтересованным лицам получить наглядное представление и принять участие в решениях, требуемых при построении классификатора для больших массивов записей.

Применение средств визуализации рассматривается при выборе объектов классификации, определении признаков и значений признаков классификации, обеспечении удобства классификатора и реализации противоречивых требований от стейкхолдеров. Среди предлагаемых решений описываются приемы использования системных классов, построения логической и физической моделей классификатора, многомерная классификация, модель данных признак-значение, логическая модель данных для описания требуемых аналитик.

В качестве предметной области рассматривается классификатор работ и услуг, приводятся примеры использования предлагаемых решений и результаты построения классификатора на крупном предприятии.

**Ключевые слова:** классификация, заинтересованное лицо, представление, признак класса, логический и физический уровни классификации, многомерная классификация, модель признак-значение, концептуальная модель данных, логическая модель данных.

## 1. Введение

Построение «правильных» классификаторов постоянно играло и продолжает играть важную роль при автоматизации предприятий, что было продемонстрировано на конференции 2023 года в РССП «Роль стандартизации в цифровизации промышленности» [1]. Для крупного предприятия наиболее важными объектами классификации являются Ресурсы, Работы/услуги и Активы.

При построении онтологической модели предприятия классификация играет важную роль для построения сначала таксономии, а затем и онтологии предприятия в целом, что было продемонстрировано на международных онтологических конференциях [2].

Однако не меньшую роль правильная классификация играет при решении практических задач автоматизации конкретных организаций, особенно крупных. При реализации системы управления основными данными предприятия важно построить удобный и эффективный инструмент работы с большим набором номенклатуры. При этом важная часть работы заключается в согласовании принимаемых решений с заинтересованными лицами – бизнес-пользователями, которые порой далеки от информационных технологий.

В соответствии с рекомендациями TOGAF [3] одна из задач архитектора информационной системы – предложить наиболее удобные представления (view) для восприятия заинтересованными лицами с самыми разными интересами. Чтобы не решать такую задачу для каждой возникающей проблемы, необходимо для каждого типового случая предложить способ подготовки такого представления (viewpoint).

## **2. Существующие способы визуализации данных**

Обзор программных средств для отображения геометрических объектов приведен в статье ученых МИФИ [4]. Обзор средств визуализации информации приведен в статье [5] и других работах сотрудников МГТУ им. Баумана. Представительная коллекция шаблонов визуализации данных представлена агентством Ferdio как результат проекта визуализации данных [6]. В статье [7] приводится общий набор применения шаблонов, в том числе для отображения иерархической классификации.

В данной работе рассматриваются возможности и результаты практического применения средств визуализации построения качественной классификации для крупного предприятия.

## **3. Существующие подходы к построению иерархической классификации**

Основные правила иерархической классификации приведены в монографии [8]. Современные эксперты по классификации используют на проектах примерно тот же набор принципов. Вариант с учетом возможностей современных систем приведен в [9] и также включает требования полноты, непротиворечивости, наличия признака классификации на каждом уровне.

## **4. Существующие проблемы классификации**

Однако на практике при построении реальных классификаторов, как правило, возникают задачи, требующие неочевидных решений. Вот некоторые из них:

- выбор объектов классификации;
- выбор признаков классификации;
- обеспечение удобства классификатора при соблюдении формальных правил;
- реализация противоречивых требований к составу классов.

Рассмотрим варианты решения этих проблем с использованием средств визуализации на примере построения классификатора работ и услуг.

## **5. Способы решения проблем классификации с использованием средств визуализации**

### **5.1. Выбор объектов классификации**

Для правильного выбора объектов классификации необходимо исходить из целей построения и дальнейшего использования классификатора и конфигурации интегрируемых информационных систем. Например, детализация услуг обычно требуется при заключении договоров с поставщиками или потребителями услуг, при подготовке спецификаций на ремонт оборудования, реке капитального строительства, при учете затрат. При наличии небольшого количества областей ситуация может быть визуализирована с использованием диаграммы Венна (Рисунок 1).

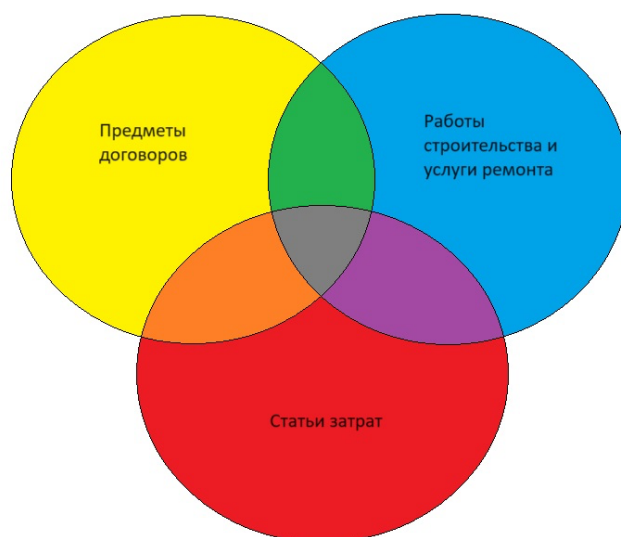


Рисунок 1. Исходные наборы записей для корпоративного справочника «Работы и услуги»

Если стоит задача интеграции трех систем соответствующего назначения, то полнота справочника может определяться зонами пересечения записей. Если же в каждой области действует несколько систем, то потребуется классификатор с охватом объединения областей. Такое решение позволит не только выполнять консолидированный анализ работ и услуг, но и обеспечивать единые корпоративные признаки учета для всего предприятия.

## 5.2. Выбор признаков классификации

Для выбора обязательных признаков классификации могут быть рассмотрены требования областей использования. Однако здесь существует опасность включить в набор признаков специфические атрибуты, присущие только данной области. Например, в бухгалтерском учете деятельность подрядчиков может списывать сразу в отчетном периоде как услуги, либо накапливаться на объектах затрат как выполняемые работы. В общем случае такая деятельность может быть и работой, и услугой и такой признак использовать нельзя.

В то же время в налоговом учете требуется формировать отчетность по ОКВЭД, который в основном определяет содержание услуг и такой признак вполне может использоваться для классификации.

Еще одна сложность заключается в том, что выбор признака необходимо сделать на каждом уровне классификации. Здесь бывает полезно посмотреть на другие справочники организации, для которых общая задача классификации всей деятельности уже решалась. Обычно на верхнем уровне используются следующие признаки классификации: подразделения, бизнес-процессы, функциональные сферы, основные сервисы и продукты.

Далее признаки будут зависеть уже от состава классифицируемых записей. Изобразить не только ветки классификации, но и значения используемых признаков позволяет Диаграмма слоев разделов (Рисунок 2) или по классификации [6] - Icicle diagram. Отличие в том, что для наших целей она развернута на 90° и дополнена важными характеристиками класса – признаками и значениями признаков вышестоящих классов.

Признак класса 2-го уровня	Значение признака	Класс 2-го уровня	Признак класса 3-го уровня	Значение признака	Класс 3-го уровня	Признак класса 4-го уровня	Значение признака	Класс 4-го уровня			
Степень участия в оформлении перевозки	Перевозка	Услуги перевозки	Объект перевозки-ки	Люди	Перевозка сотрудников						
				Документы	Доставка документов						
				Грузы	Перевозка грузов	Вид транспорта	Автомобиль	Перевозка автомобильным транспортом			
							Ж/д	Перевозка железной дорогой			
							Прочие	Перевозка прочими видами транспорта			
	Оформление перевозки	Услуги, не связанные с перевозкой	Вид работ	Оформление	Услуги по оформлению						
				Перевалка	Услуги по перевалке						
				Сервисное обслуживание	Услуги по сервису транспортных средств						
	Комплексные услуги	Экспедиторские услуги									

Рисунок 2. Использование диаграммы слоев разделов для демонстрации признаков и их значений

### 5.3. Обеспечение удобства классификатора при соблюдении формальных правил

Одна из распространенных ошибок при построении классификаторов – стремление во чтобы то ни стало «выровнять» количество уровней и длину веток классификации. Критерием «удобства» при этом выбирается примерно одинаковая сложность каждого уровня.

На первый взгляд, таким классификатором удобнее пользоваться, поскольку визуально он выглядит более стройным. Однако это обманчивое впечатление. При таком подходе нарушается принцип выделения явного признака классификации на каждом уровне. Сложные предметные области, как правило, не симметричны. В части функционального моделирования подобное наблюдение было описано в методике по функциональному моделированию [10], однако это правило применимо и для классифика-

ции данных. Более того, если Вы видите «красивую» симметричную классификацию, то уже визуально понятно, что она, скорее всего, не верна.

С другой стороны, если строго соблюдать правило один класс - один признак, то на практике мы можем получить логически правильно построенный, но абсолютно неудобный для использования классификатор.

Выходом здесь является разделение построения классификационного дерева на два этапа: построение сначала логической модели, а затем физической модели для итоговой реализации в системе, по аналогии с подходом к моделированию данных [11]. На логическом уровне полезно зафиксировать явно все возможные признаки и строгое деление на каждом уровне. На физическом уровне можно аккуратно объединить наиболее разветвленные участки с учетом знаний предыдущего уровня, перенося сокращаемые признаки в название класса.

На Рисунке 3 при использовании стандартного дерева мы видим, что можем укоротить третью ветку классификации, объединив признаки «Объект перевозки» и «Вид транспорта».

#### **5.4. Реализация противоречивых требований к составу классов**

На большом предприятии возможны разные требования к требуемым классам. Например, в закупочном подразделении (или организации холдинга) важно учитывать перевозку грузов по видам транспорта, чтобы выбирать наиболее эффективные способы. А для сбытового подразделения/организации может быть важно делить услуги перевозки грузов по видам продукции для более простого расчета затрат.

В принципе такая задача решается декартовым произведением требуемых классов. Такой подход допустим внутри одной организации или для небольших объемов. Однако в рамках холдинга каждой организации может потребоваться свой признак классификации на том же уровне обобщения.

Такое ограничение также имеет способ решения в рамках единственного иерархического классификатора. Он заключается в использовании системных признаков классификации, где в качестве значения признака выступает признак класса следующего уровня. В нашем примере признаком будет являться «Признак классификации», а значениями – «Вид транспорта» и «Вид груза» (Рисунок 4).

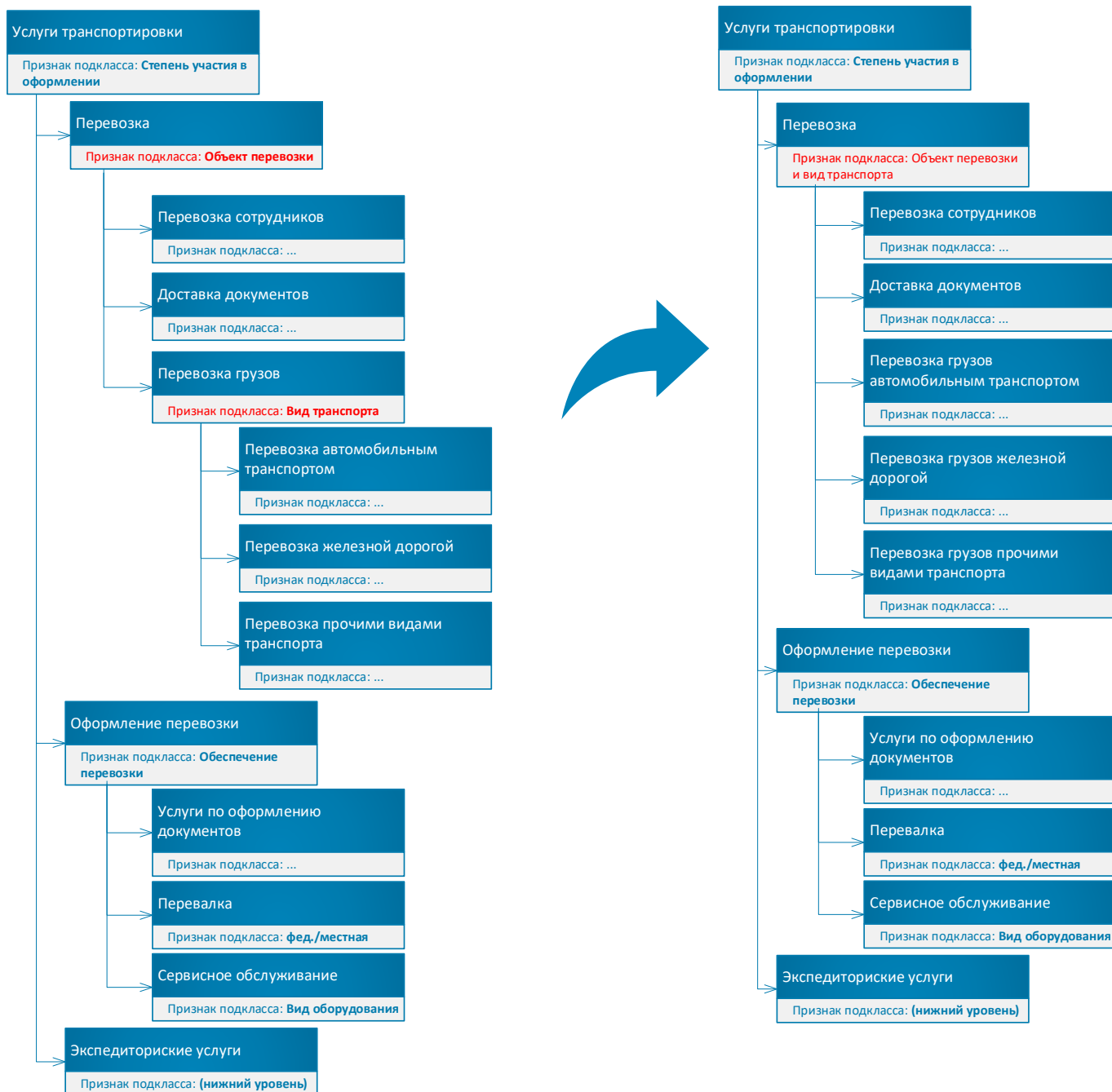


Рисунок 3. Сокращение длины веток классификации на диаграмме «Дерево»

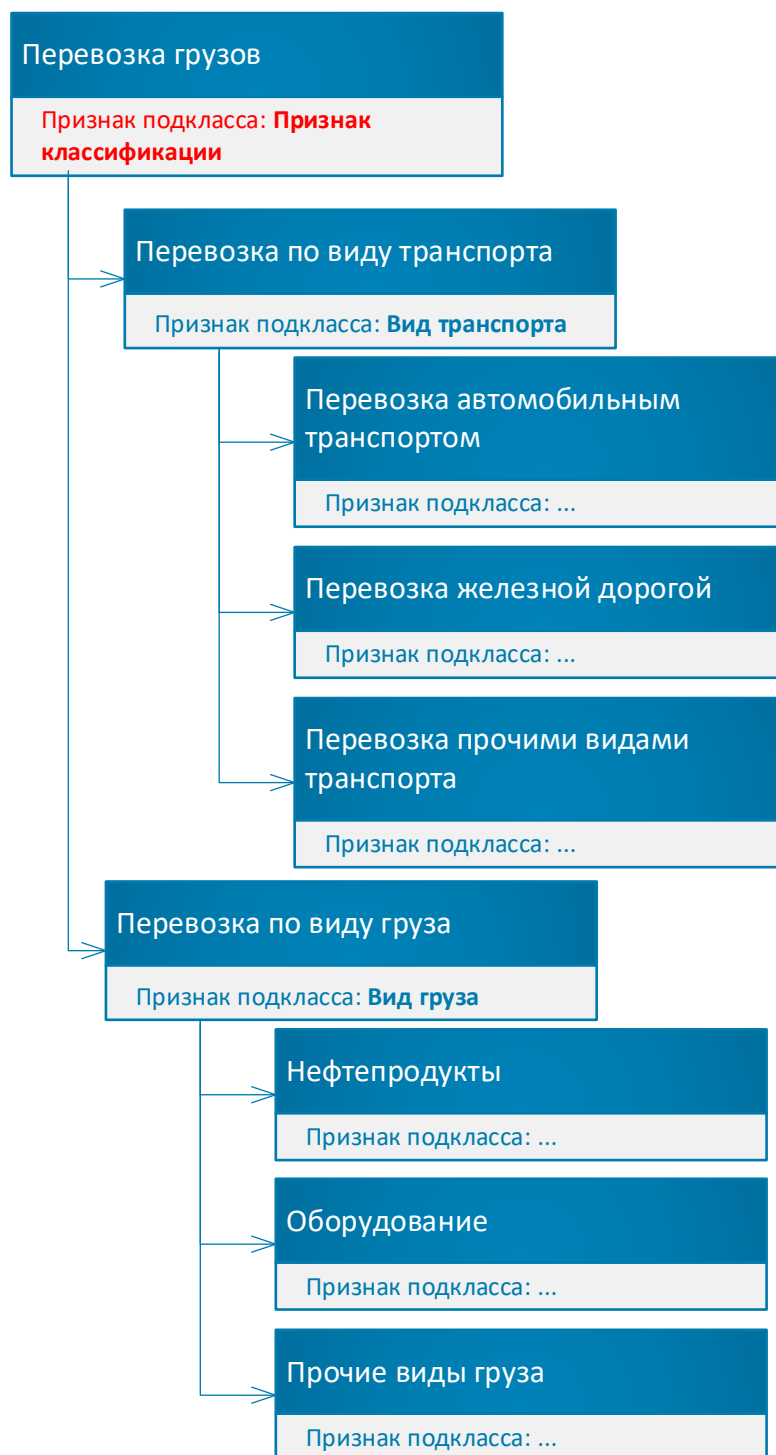


Рисунок 4. Использование системного признака класса

Следует отметить, что такой способ применим только в случае, если

- 1) классы ниже данного уровня используются только локально, все корпоративные признаки могут быть привязаны на уровнях выше.
- 2) использование параллельных классификаций существенно менее эффективно: требует дополнительных разработок в смежных системах, усложняет модели данных, и т.д.

Более перспективным является выход за пределы использования единой иерархии. Для поиска наиболее эффективных вариантов решения могут помочь более сложные модели данных, позволяющие отразить необходимые сущности и связи. Ниже будут показаны три способа решения, для представления которых используются приемы моделирования данных. Среди множества возможных нотаций [12], [13] наиболее удоб-

ной зарекомендовала себя визуализация отношений с использованием нотации Мартина (другое название - Crow's Foot, «Воронья лапка»), более наглядно отражающая кардинальность связываемых сущностей и компактно представляющая набор атрибутов.

### 3.1) Многомерная классификация

В случае, если одни и те же объекты должны быть проклассифицированы с использованием разных признаков классификации, то возможно, подойдет вариант перехода к фасетной классификации. Концептуальная модель данных, показывающая данный способ решения приведена на Рисунке 5.

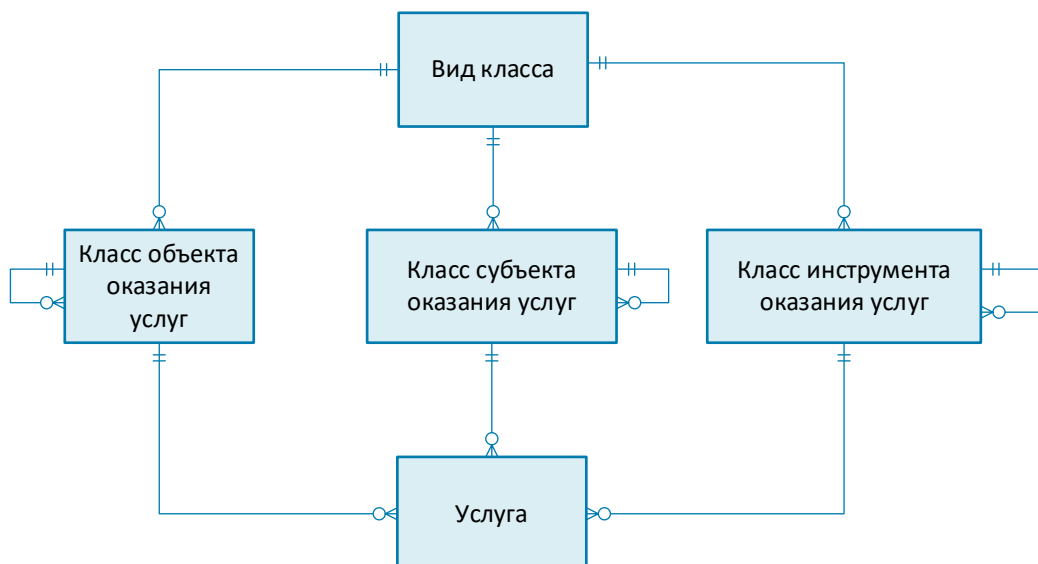


Рисунок 5. Концептуальная модель данных многомерной классификации услуг

Многомерная классификация открывает дополнительные возможности для визуального анализа данных. На рисунках 5 и 6 приведены примеры диаграмм при использовании многомерной классификации услуг.



Рисунок 6. Объемная гистограмма



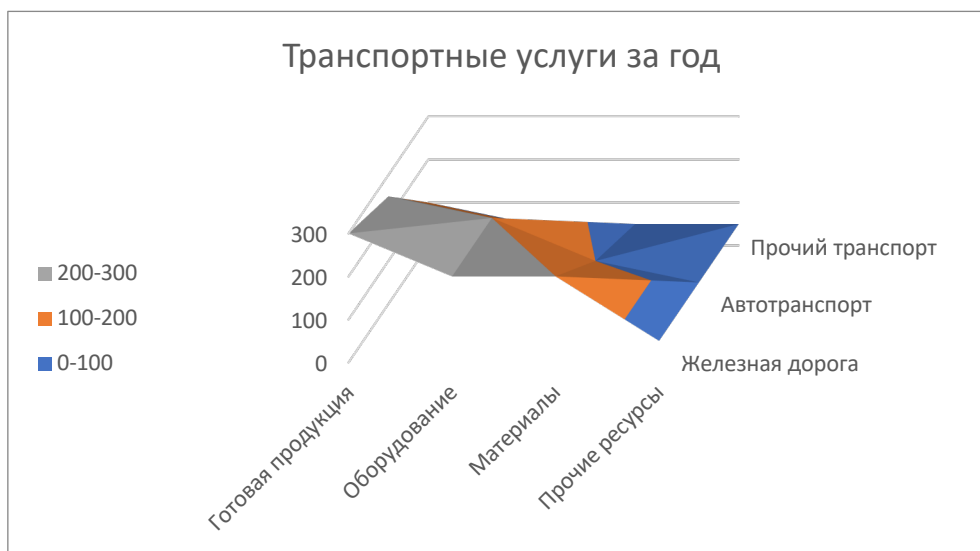


Рисунок 7. Поверхностная диаграмма

3.2) В случае, если набор дополнительных признаков для каждого подкласса уникален и содержит большое количество линейных значений, можем воспользоваться другим подходом, подразумевающим хранение своих наборов признаков для каждого класса (Рисунок 6). Следует отметить, что наиболее эффективен такой подход для классификации материалов. Подробнее о решении проблем автоматизации справочника материалов см. в [14], [15].



Рисунок 6. Концептуальная модель данных использования выделенных наборов значений признаков для класса

### 3.3) Использование отдельных аналитик

Для иллюстрации данного метода более пригодна уже логическая модель данных (Рисунок 7). В данном случае она показывает, что для получения дополнительной аналитике (вид груза) аналитик может подойти классификатор товаров, уже используемые при оформлении транзакционных документов перевозки.



Рисунок 7. Логическая модель данных связи услуг и классов грузов

## 6. Заключение

В данной работе предложены варианты решения задач классификации работ и услуг, которые успешно опробованы на проектах развития нормативно-справочной информации в крупной российской компании с участием студентов МИФИ [16], [17]. В качестве инструмента классификации и в целом управления справочными данными использовалась корпоративная система управления НСИ на платформе IBM Product Master 12.0, обладающая гибкими средствами управления структурой справочных данных.

Среди средств управления классификациями можно отметить следующие:

- Удобные средства визуализации, в том числе для многомерных классификаций
- Возможность импорта стандартных или определения пользовательских классификаций
- Просмотр и изменение записей с использованием различных классификаций
- Управление маппингом классов из разных иерархий

Среди дополнительных средств визуализации можно отметить следующие:

- Отображение настроенных рабочих потоков согласования записей различными категориями пользователей в виде диаграмм последовательностей;
- Отображение потоков обработки справочных данных при настройке дополнительных процедур проверки качества данных;
- Подключение внешних средств визуализации, например, карт для показа местоположения объекта справочника с заданными координатами.

Построенный классификатор работ и услуг имеет порядка 400 классов с заданными признаками классификации, используется для классификации более 6000 исходных записей информационных систем. В качестве источников данных использовались 6 предметных областей (Закупки, Платежи, Затраты, Ремонты, Инвестиции, Бухгалтерский учет). Справочник работ и услуг централизованно ведётся более чем для 10 информационных систем холдинга.

## Список литературы

1. Кукшев В.И., «Перспективы разработки классификаторов промышленной продукции в ЕАЭС», 12 12 2023. [В Интернете]. Available: <http://www.rgtr.ru/data/events/2023/SMART%2C%2012.12.2023/8.%20%D0%9A%D1%83%D0%BA%D1%88%D0%B5%D0%B2%20%D0%92.%D0%98..pdf>.

2. H. Hedden, «Turning a Taxonomy into an Ontology,» 17 11 2021. [В Интернете]. Available: <https://www.hedden-information.com/wp-content/uploads/2021/11/Turning-a-Taxonomy-into-an-Ontology.pdf>.
3. Open Group, «TOGAF,» Open Group, April 2022. [В Интернете]. Available: <https://pubs.opengroup.org/togaf-standard/index.html>.
4. М.Н.Стриханов, Н.Н.Дегтяренко, В.В.Пилюгин, Е.Е. Маликова, Н.А. Матвеева, В.Д. Аджтев, А.А.Пасько, «Опыт компьютерной визуализации наноструктур в НИЯУ МИФИ,» Scientific Visualization, т. 1, № 1, pp. 1-18, 2009.
5. О.В.Пескова, «О визуализации информации,» Вестник МГТУ им. Н.Э. Баумана. Сер. «Приборостроение», 2012.
6. Ferdio, « Data Viz Project,» Ferdio, [В Интернете]. Available: <https://datavizproject.com>. [Дата обращения: 30 07 2024].
7. И.К. Романова, «СОВРЕМЕННЫЕ МЕТОДЫ ВИЗУАЛИЗАЦИИ МНОГОМЕРНЫХ ДАННЫХ: АНАЛИЗ, КЛАССИФИКАЦИЯ, РЕАЛИЗАЦИЯ, ПРИЛОЖЕНИЯ В ТЕХНИЧЕСКИХ СИСТЕМАХ,» Наука и образование: научное издание МГТУ им. Н.Э. Баумана, № 3, pp. 133-167, 2016.
8. Айвазян С.А., Бежаева З.И., Староверов О.В., Классификация многомерных наблюдений, Москва: Статистика, 1974.
9. «Требования к иерархической системе классификации,» [В Интернете]. Available: <https://studylib.ru/doc/4675234/trebovaniya-k-ierarhicheskoy-sisteme-klassifikacii>. [Дата обращения: 30 07 2024].
10. К. М. Давид Марка, Методология структурного анализа и проектирования: Пер. с англ., Москва: Оригинал-макет: «Мета-технология», Полиграфия: ГМП «Первая Образцовая типография», 1993.
11. Mark Mosley, Michael Brackett, Susan Earley, Deborah Henderson, The DAMA Guide to the Data Management Body of Knowledge, NJ USA: DAMA International, 2010.
12. М.Нижельская, Н.Класен, А.Павлова, «Сущности и связи: как и для чего системные аналитики создают ER диаграммы,» 17 05 2023. [В Интернете]. Available: <https://practicum.yandex.ru/blog/chto-takoe-er-diagramma/>.
13. «Нотации модели сущность-связь (ER диаграммы),» 11 09 2021. [В Интернете]. Available: <https://pro-prof.com/archives/8126>.
14. Дзенгелевский А.Е., Могилат А.С., «Классификаторы КССС: теперь и готовая продукция,» ITime – Информационные технологии в ТЭК, т. 2(12), pp. 26-29, 2010.
15. Дзенгелевский А.Е., «Уровни обобщения при учете товарно-материальных ценностей,» Актуальные проблемы гуманитарных и естественных наук, т. 12 (59), № I, pp. 82-90, 2013.
16. С. А. Дзенгелевский А.Е., «История КССС: Единая корпоративная система словарей и справочников,» ITime – Информационные технологии в ТЭК, 2007.
17. Х. Ш. Дзенгелевский А.Е., «Опыт создания и развития корпоративной системы управления нормативно-справочной информацией,» Научно-методический журнал "Межотраслевая информационная служба», т. 2(12), pp. 22-26, 2012.

## Data Classification with Using Visualization Tools

Andrey Dzengelewski<sup>1</sup>

National Research Nuclear University MEPhI (Moscow Engineering Physics Institute),  
Moscow, Russia

<sup>1</sup> ORCID: 0009-0002-4770-1421, [Dzengelewski@gmail.com](mailto:Dzengelewski@gmail.com)

## Abstract

This article discusses ways to use visualization tools to build object classifiers during automation of a large enterprise. The proposed approaches allow stakeholders to get a visual representation and participate in the decisions required when building a classifier for large arrays of records.

The use of visualization tools is considered when selecting classification objects, determining the attributes and values of classification attributes, ensuring the convenience of the classifier and implementing conflicting requirements from stakeholders. Among the proposed solutions, the methods of using system classes, building logical and physical models of the classifier, multidimensional classification, attribute-value data model, logical data model for describing the required analytics are described.

The subject area is a classifier of works and services, examples of using the proposed solutions and the results of building a classifier at a large enterprise are given.

**Keywords:** classification, stakeholder, view, class attribute, logical and physical levels of classification, multidimensional classification, attribute-value model, conceptual data model, logical data model.

## References

1. Kukshev V.I., "Prospects for the Development of Industrial Product Classifiers in the EAEU," 12 12 2023. [Online]. Available: <http://www.rgtr.ru/data/events/2023/SMART%2C%2012.12.2023/8.%20%Do%9A%D1%83%Do%BA%D1%88%Do%B5%Do%B2%20%Do%92.%Do%98..pdf>.
2. H. Hedden, "Turning a Taxonomy into an Ontology," 17 11 2021. [Online]. Available: <https://www.hedden-information.com/wp-content/uploads/2021/11/Turning-a-Taxonomy-into-an-Ontology.pdf>.
3. The Open Group, "TOGAF," The Open Group, April 2022. [Online]. Available: <https://pubs.opengroup.org/togaf-standard/index.html>.
4. M.N.Strikhanov, N.N.Degtyarenko, V.V.Pilyugin, E.E. Malikova, N.A. Matveeva, V.D.Adzhtev, A.A.Pasko, "Experience of computer visualization of nanostructures at MEPhI," Scientific Visualization, vol. 1, no. 1, pp. 1-18, 2009.
5. O.V. Peskova, "On information visualization," Bulletin of Bauman Moscow State Technical University. Series "Instrument Engineering", 2012.
6. Ferdio, "Data Viz Project," Ferdio, [ In Internet ]. Available: <https://datavizproject.com>. [Accessed: 30 07 2024].
7. I.K. Romanova, "MODERN METHODS OF MULTIDIMENSIONAL DATA VISUALIZATION: ANALYSIS, CLASSIFICATION, IMPLEMENTATION, APPLICATIONS IN TECHNICAL SYSTEMS," Science and Education: scientific publication of Bauman Moscow State Technical University, No. 3, pp. 133-167, 2016.
8. Ayvazyan S.A., Bezhaeva Z.I., Staroverov O.V., Classification of multivariate observations, Moscow: Statistics, 1974.

9. "Requirements for a hierarchical classification system," [Online]. Available: <https://studylib.ru/doc/4675234/trebovaniya-k-ierarhicheskoj-sisteme-klassifikacii>. [Accessed: 30 07 2024].
10. K. M. David Mark, Methodology of structural analysis and design: Translated from English, Moscow: Original layout: "Meta-technology", Printing: GMP "First Model Printing House", 1993.
11. Mark Mosley, Michael Brackett, Susan Earley, Deborah Henderson, The DAMA Guide to the Data Management Body of Knowledge, NJ USA: DAMA International, 2010.
12. M. Nizhelskaya, N. Klassen, A. Pavlova, "Entities and Relationships: How and Why System Analysts Create ER Diagrams," 17 05 2023. [Online]. Available: <https://practicum.yandex.ru/blog/chto-takoe-er-diagramma/>.
13. "Entity-Relationship Model Notations (ER Diagrams)," 11 09 2021. [Online]. Available: <https://pro-prof.com/archives/8126>.
14. Dzengelevsky A.E., Mogilat A.S., "Classifiers of the KSSS: now also finished products," ITime – Information technologies in the fuel and energy complex, v. 2(12), pp. 26-29, 2010.
15. Dzengelevsky A.E., "Levels of Generalization in Accounting for Inventory," Actual Problems of Humanities and Natural Sciences, Vol. 12 (59), No. I, pp. 82-90, 2013.
16. S. A. Dzengelevsky A. E., "History of the KSSS: Unified Corporate System of Dictionaries and Reference Books," ITime – Information Technologies in the Fuel and Energy Complex, 2007.
17. H. Sh. Dzengelevsky A.E., "Experience in the creation and development of a corporate system for managing normative and reference information," Scientific and Methodological Journal "Inter-Industry Information Service", v. 2(12), pp. 22-26, 2012.