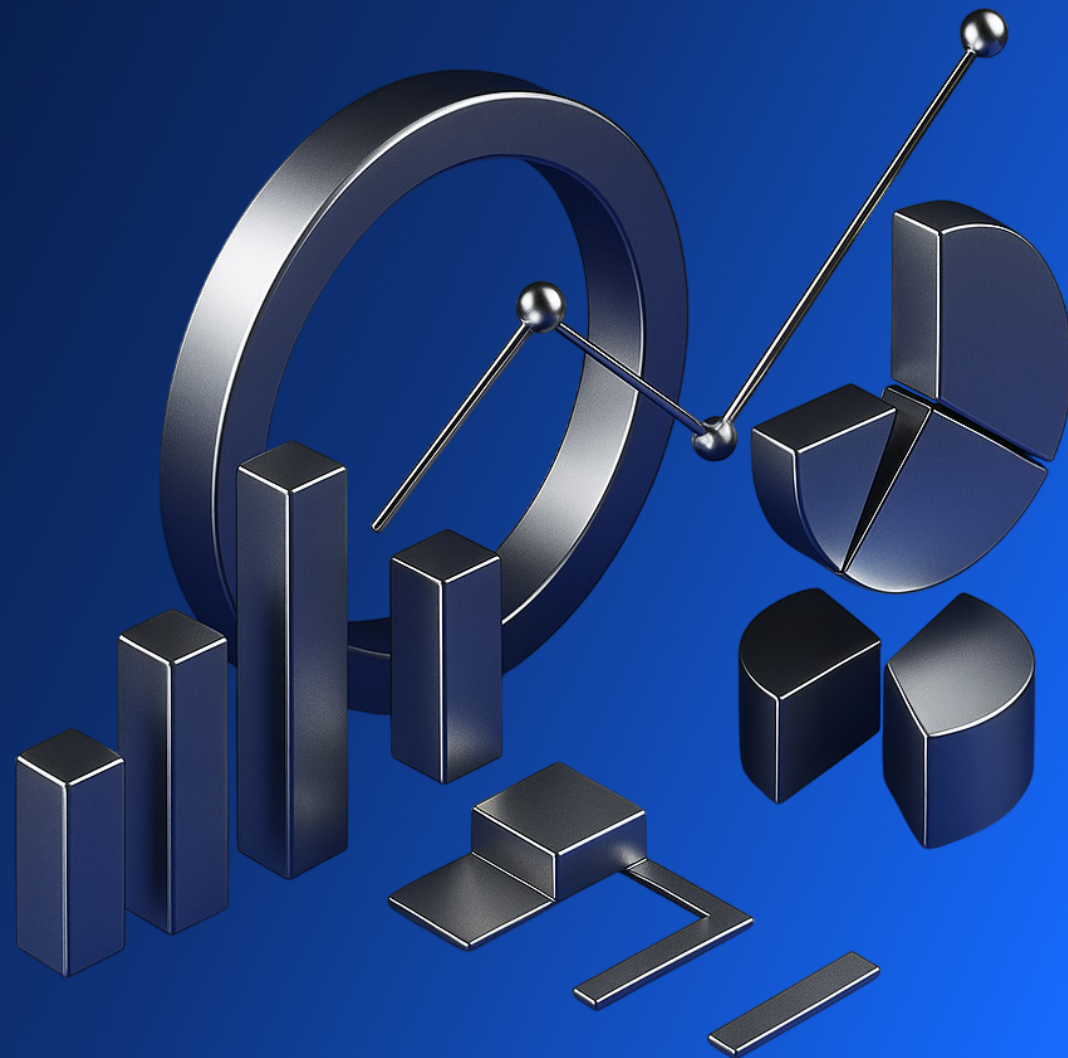


Data Science на базе Loginom

опыт ООО «Газпром проектирование»

Евгения Клевачкина

Вадим Савинков



ИССЛЕДОВАНИЯ И ОБОСНОВАНИЕ



- Преинвестиционные исследования, разработка ТЭО и инвестиционных замыслов
- Научно-исследовательские и опытно-конструкторские работы (НИОКР)

РАЗРАБОТКА



- Технологии подготовки, очистки и переработки природного газа
- Нормативная, техническая и ценовая документация

ПРОЕКТИРОВАНИЕ И ПРОИЗВОДСТВО



- Комплексные проектно-изыскательские работы
- Проектирование и изготовление нефтегазового оборудования

ГАЗИФИКАЦИЯ



- Развитие газоснабжения для потребителей в РФ и иностранных партнеров

10 лет на рынке

Генеральный проектировщик
инвестиционных проектов
Группы Газпром



Управление корпоративными данными



Корпоративные данные

стратегический актив Общества

Нормативно-справочная информация


основа корпоративных данных

В целях обеспечения единообразного учета МТР поставлена задача, включающая:

- 1 Обязательное заполнение кода ОКПД2 для новых МТР
- 2 Проверку заполнения кода ОКПД2 для существующих МТР



Разработанная в Обществе ИУС ЕБД МТР содержит порядка 400 000 записей МТР



Общество с ограниченной ответственностью «Газпром проектирование»
(ООО «Газпром проектирование»)

ПРИКАЗ

«15» апреля 2025 г. Санкт-Петербург № 224


Об организации и обеспечении учета МТР в ИУС ЕБД МТР в соответствии с Общероссийским классификатором продукции по видам экономической деятельности и Классификатором строительных ресурсов Российской Федерации

В целях обеспечения единообразного учета материально-технических ресурсов в компаниях Группы Газпром (далее – МТР) и возможности в последующем их хранения и отображения в ООО «Газпром проектирование» в соответствии с базовым и прикладным классификаторами Единого справочника МТР ПАО «Газпром» (далее – ЕС МТР), разрабатываемого в рамках проекта развития ОБД НСИ ПАО «Газпром» в части подсистемы ведения единого справочника МТР ПАО «Газпром» и компаний Группы Газпром, осуществления возможности обмена данными в автоматизированном режиме и синхронизации данных по применяемым в проектных решениях МТР на основании кодов Общероссийского классификатора продукции по видам экономической деятельности (далее – ОКПД2) и кодов Классификатора строительных ресурсов (далее – КСР), принимая во внимание техническую готовность ИУС ЕБД МТР по учету в записях МТР кодов указанных классификаторов

ПРИКАЗЫВАЮ:

1. Заместителю главного инженера по технической политике (В.С. Сидоров) организовать работу и силами вверенных в управление подразделений с даты регистрации настоящего приказа обеспечить мониторинг Общероссийского классификатора продукции по видам экономической деятельности и Классификатора строительных ресурсов на предмет их актуализации и начиная с 01.06.2025 передачу их редакций в адрес ЦИТ Саратовского филиала Общества в целях размещения в ИУС ЕБД МТР по мере их обновления.

2. Заместителю генерального директора по информационным технологиям (В.В. Гурьянов) организовать работу и силами вверенных в управление подразделений обеспечить:



2 000016 311990

1**Дубли позиций**

разные идентификаторы у полностью идентичных МТР

2**Отсутствие кода ОКПД2**

для существующих записей МТР

3**Неполнота данных**

значения ОКПД2 заполнены до 2 и 3 уровней

207 000 позиций

размер выборки текущего исследования

33% позиций

не содержат значений кодов ОКПД2

14% позиций

содержат некорректные значения кода ОКПД2



Снизить трудозатраты на ввод данных работниками НСИ на 50% и повысить качество корпоративных данных о МТР на 15 % в части ОКПД2



Основные метрики

- Доля пропущенных значений ОКПД2 в записях МТР
- Доля дубликатов МТР
- Трудозатраты обработки одной записи



Ручное заполнение пропусков

работники отделов
сопровождения НСИ/
проектных
подразделений



Программирование

SQL

Jupyter Notebook: Python



Loginom

выбранное решение



Имеющийся опыт

Коммерческий Data Science:
EDA, реализация ETL-
пайплайнов, ML

Технологический стек:
Python, SQL, Visiology



Формат обучения

Loginom Skills



Сообщество Loginom




@LOGINOM



Гипотеза

Записи МТР с отсутствующим значением кода классификатора ОКПД2 могут быть идентифицированы как дубликаты или аналогичные записи с заполненным кодом ОКПД2 по другим атрибутам, схожим по смыслу



Решение



Сценарий Loginom

1

Отбор наиболее значимых признаков

2

Предобработка данных

3

Разделение на 2 набора (с заполненными и незаполненными ОКПД2)

4

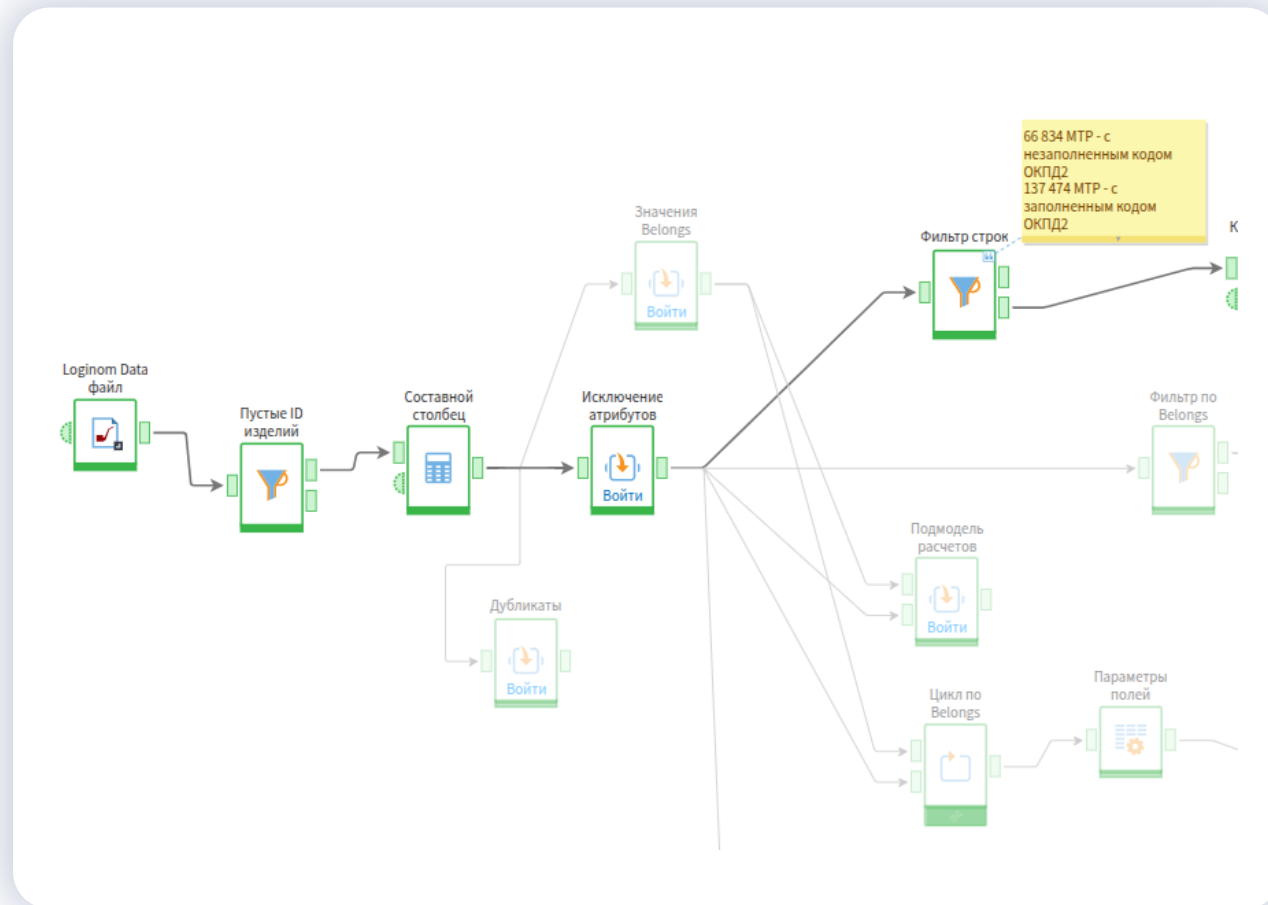
Сопоставление изделий

5

Расчет расстояния Дамерау-Левенштейна

6

Отбор значений



Расчет расстояния ДЛ

Расстояние Дамерау-Левенштейна

мера разницы двух строк символов, определяемая как минимальное количество операций вставки, удаления, замены и транспозиции (перестановки двух соседних символов), необходимых для перевода одной строки в другую.

Реализация Loginom
Калькулятор



DamLevDist(Строка1, Строка2)

$$d_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{если } \min(i,j) = 0, \\ \min \begin{cases} d_{a,b}(i-1,j) + 1 \\ d_{a,b}(i,j-1) + 1 \\ d_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{если } i,j > 1 \text{ и } a_i = b_{j-1} \text{ и } a_{i-1} = b_j \\ \min \begin{cases} d_{a,b}(i-1,j) + 1 \\ d_{a,b}(i,j-1) + 1 \\ d_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{иначе,} \end{cases}$$

#	ab Код ОКПД2 +	ab Наименование и техническая характеристика оборудова...	12 ид мтр -	12 Расстояние ДЛ
1	22.19.73.110	Пластина 2Ф-I-ТМКЩ-C-2	77 076	2
2	22.19.73.110	Пластина 2Ф-I-ТМКЩ-C-8	84 642	2
3	22.19.20.120	Пластина 1Н-I-ТМКЩ-C2-2	165 718	2
4	22.19.73.110	Пластина 1Ф-I-ТМКЩ-C-1	205 750	2
5	22.19.73.110	Пластина 2Ф-I-ТМКЩ-C2-1	269 281	2
6	22.19.20.120	Пластина 2Ф-I-ТМКЩ-C2-3	269 596	2
7	22.19.73.110	Пластина 1Ф-I-ТМКЩ-C2-4 250x250	271 744	2
8	22.19.20.120	Пластина 1Ф-I-ТМКЩ-C-2	451 611	2
9	25.91.11.000	Баллон для азота	344 253	2
10	28.99.39.190	Головка рукавная напорная Ру 1,6 МПа	66 867	2
11	28.99.39.190	Арматура фонтанная	306 396	3
12	22.19.73.110	Пластина 2Н-I-ТМКЩ-C1-2	151 793	3
13	22.19.73.110	Пластина 2Н-I-ТМКЩ-C1-3	151 794	3
14	22.19.20.120	Пластина 2Н-I-ТМКЩ-C2-2	162 347	3
15	22.19.20.120	Пластина 2Н-I-ТМКЩ-C2-3	162 351	3
16	28.99.39.190	Комплекс термического обезвреживания жидких отходов	431 047	3
17	22.19.30.136	Рукав высокого давления с накидной гайкой, в комплекте с пере...	430 663	4
18	25.91.11.000	Баллон для азота	71 918	4
19	25.91.11.000	Баллон-1И-азот-40-9,8-НЛ-219	405 024	4
20	28.29.22.110	Огнетушитель углекислотный, переносной. Масса заряда 3кг. Га...	34 047	4
21	28.99.39.190	Комплекс термического обезвреживания жидких отходов	332 180	4
22	32.50.13.190	Спринцовка пластизольная (35 мл)	142 102	5
31	22.19.71.190	Спринцовка резиновая (35 мл)	142 106	5


Таблица Форма

Сортировка



Результаты

По итогам проверки полученных результатов гипотеза оказалась подтверждена. Записи МТР с пропусками кода ОКПД2 идентифицируются как дубликаты/аналоги по атрибутам. Совокупное время выполнения сценария составило чуть менее 3 ч.





Очередной шаг в направлении цифровой трансформации

переход от администрирования корпоративных данных к оказанию сервисной функции по их управлению



Экономические эффекты

- Коды ОКПД2 присвоены для 36 000 позиций
- Выявлено более 800 дублей
- Доля заполненных записей возросла до 70%
- Сэкономлено около 4000 часов рабочего времени
- Время обработки одной записи составляет 3,3 сек

Решение рутинных задач средствами Loginom

НМТР

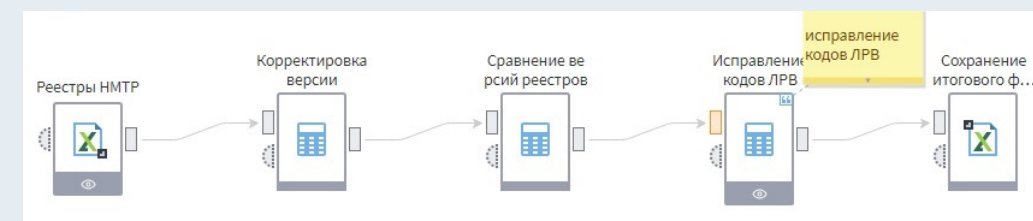
невыпущенные материально-технические ресурсы

1 Разрозненные источники данных НМТР

2 Отсутствие единых правил ведения реестра НМТР

Сценарий Loginom

для исправления кодов лимитно-разделительной ведомости и сравнения версий реестров



```
Предпросмотр... | && || ! | == != < > <= >= | 9.0 " " 31 ⌚ false true  
1 if (Kod_LRV_ === "Без привязки") {return Kod_LRV_}  
2  
3 const delimiters = ["\"", ";", " ", ",", "\n", "\t"];  
4 let cleaned = Kod_LRV_.replace(/\\n/g, " ").trim();  
5 let parts = cleaned.split(new RegExp(delimiters.join('|')));  
6 parts = parts.filter(part => part.trim().length > 0);  
7 parts = parts.map(part => part.replace(/\\s+/g, "").trim());  
8 return parts.join("; ") + ";";
```

Контакты



Евгения Клевачкина



eklevachkina@proektirovanie.gazprom.ru



Вадим Савинков



vrsavinkov@proektirovanie.gazprom.ru