

Разработка систем индексации данных для производственно-экономического и трудового секторов пенитенциарной системы

Д.С. Пономарев

*Научно-исследовательский институт Федеральной службы исполнения наказаний,
г. Москва*

*Ижевский государственный технический университет имени М.Т. Калашикова,
г. Ижевск*

Аннотация: Разработка систем бизнес-аналитики, принятия решений и планирования ресурсов является одной из важнейших составляющих практически любого предприятия. В данном случае, предприятия и производства пенитенциарной системы не являются исключением. В работе рассмотрена проблема взаимосвязи существующих баз данных и статистических отчетных форм производственно-экономического и трудового секторов пенитенциарной системы. Установлено, что косвенно-взаимосвязанные параметры достаточно трудно сопоставить из-за разных систем учета данных, а также утвержденных статистических форм. Одной из первых ступеней в решении поставленной проблемы может стать введение обобщенной системы индексации данных. В работе рассмотрены системы индексации данных, построение их иерархических структур, а также возможность практического применения с использованием *SQL*. Рассмотрены примеры реализации при помощи технологии *ORM* и языка *Python*.

Ключевые слова: базы данных, индексация, *ORM*, *SQL*, *Python*, производственный сектор, экономические показатели, пенитенциарная система.

Введение

На сегодняшний день учет статистических данных является важной задачей для практически любого крупного производственного предприятия. И производственные предприятия пенитенциарной системы в данном вопросе не являются исключением [1]. Для производственно-экономического сектора пенитенциарной системы сбор и учет информации закреплен утвержденными ведомственными статистическими формами (Приказ ФСИН России №754 от 30.08.2019). Помимо учета статистической информации для производственно-экономического сектора, в уголовно-исполнительной системе существуют и другие формы учета данных, которые затрагивают непосредственно спецконтингент: показатели медицинского обеспечения, социальных показателей и общей характеристики спецконтингента (учет данных показателей ведется согласно Приказу ФСИН России №1033 от

12.11.2019; Приказу ФСИН России №661 от 15.09.2020; Приказу ФСИН России от 01.08.2014 № 398).

Первая проблема здесь состоит в том, что учтенные данные в целом недостаточно изучены: на сегодняшний день отсутствуют какие-либо утвержденные и внедренные методы и методики прогноза, поддержки принятия решений, автоматизации производственной деятельности в пенитенциарной системе на основе полученных знаний. Таким образом, собранные данные представляют интерес как с практических позиций организации и поддержания производственно-трудовой и экономической деятельности, так и с научной точки зрения. Однако, здесь стоит отметить, что данные разработки достаточно активно ведутся в научно-исследовательских подразделениях пенитенциарной системы [2, 3] в сотрудничестве с территориальными органами и медико-санитарными частями [4]. Второй проблемой является отсутствие каких-либо попыток связать данные между разными статистическими формами. Например, отсутствуют какие-либо научные обоснования взаимосвязи характеристик трудоустроенного спецконтингента и показателей медицинского обеспечения, социальных показателей и общей характеристикой спецконтингента пенитенциарных учреждений.

Таким образом, представляется актуальным разработка систем и алгоритмов сопоставления баз данных. Актуальным также представляется создание новых баз данных, использование которых будет направлено непосредственно на решение проблем прогноза, поддержки принятия решений и автоматизации производственной деятельности. Применение основ реляционной теории и разработки систем индексации [5, 6] может явиться первой ступенью в решении поставленных задач. Рассмотрим более подробно возможную систему индексации данных и ее структуры.

Разработка системы индексации

Для индексации данных были определены группы учета: X – данные, отражающие значения параметров и показателей согласно приведенным ранее Приказам; Y – срезы, которые можно осуществить относительно X ; временной срез учитывается отдельно и обозначен t .

Для X было принято обозначение: $X_{m,p,o}$, где m – индекс номера формы учета; p – индекс номера раздела в форме учета данных; o – индекс номера параметра в разделе. Для $m=1$ будет соответствовать форма ОТАО-1 (то есть отчеты о трудовой адаптации); $m=2$ соответственно форма ОТАО-2 (отчеты по производственной и экономической деятельности); $m=3$ – форма ФСИН-1 (что соответствует статистическим отчетам по итогам деятельности учреждений); $m=4$ – соответствует форме ФСИН-6, которая отражает сведения о социально-значимых заболеваниях. Исходя из m , далее можно выбрать один из разделов (что будет соответствовать индексу p). Для $m=1$ будет справедливо: $p=1$ – сведения о привлечении спецконтингента к трудовой адаптации; $p=2$ – заработная плата спецконтингента; $p=3$ – возмещение причиненного ущерба; $p=4$ – нормирование труда; $p=5$ – сведения о наличии и составе спецконтингента (трудовых ресурсов); $p=6$ – сведения о фактической численности и фонде заработной платы производственного персонала; $p=7$ – сведения о трудовой адаптации спецконтингента в отраслевом разрезе. Для $m=2$ значения p будут: $p=1$ – производственная и экономическая деятельность учреждений уголовно-исполнительной системы; $p=2$ – производственная и экономическая деятельность в отраслевом разрезе. Для $m=3$: $p=1$ – сведения о количестве учреждений, исполняющих уголовные наказания в виде лишения свободы, СИЗО и характеристике содержащихся в них лицах. Для $m=4$: $p=1$ – сведения о социально значимых заболеваниях у лиц, содержащихся в учреждениях уголовно-исполнительной системы Российской Федерации; $p=2$ – сведения

об отдельных показателях деятельности медицинской службы. Таким образом, будут получены следующие условия (1-4). Для X при $m=1$ будет справедливо условие (1):

$$X_{m=1} \Rightarrow \begin{cases} p = 1 \Rightarrow o \in \{1,2,\dots,21\} \\ p = 2 \Rightarrow o \in \{1,2,\dots,40\} \\ p = 3 \Rightarrow o \in \{1,2,\dots,36\} \\ p = 4 \Rightarrow o \in \{1,2,\dots,12\}, \\ p = 5 \Rightarrow o \in \{1,2,\dots,23\} \\ p = 6 \Rightarrow o \in \{1,2,\dots,8\} \\ p = 7 \Rightarrow o \in \{1,2,\dots,24\} \end{cases} \quad (1)$$

для X при $m=2$ (2):

$$X_{m=2} \Rightarrow \begin{cases} p = 1 \Rightarrow o \in \{1,2,\dots,61\} \\ p = 2 \Rightarrow o \in \{1,2,\dots,37\} \end{cases}, \quad (2)$$

для X при $m=3$ (3):

$$X_{m=3} \Rightarrow p = 1, o \in \{1,2,\dots,159\}, \quad (3)$$

для X при $m=4$ (4):

$$X_{m=4} \Rightarrow \begin{cases} p = 1 \Rightarrow o \in \{1,2,\dots,81\} \\ p = 2 \Rightarrow o \in \{1,2,\dots,48\} \end{cases}. \quad (4)$$

Представленная система индексации для X имеет иерархическую структуру (рисунок 1).

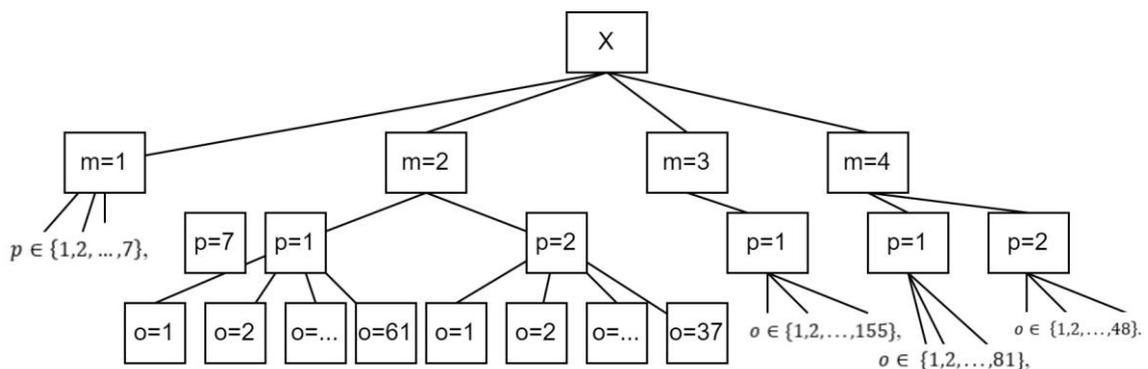


Рис. 1. – Структура индексов для исследуемых параметров

Аналогичным образом введем индексацию для срезов по представленным параметрам и обозначим ее $Y_{u,v}$: где u – срез по учреждениям ($u=1, 3, 4$) или отраслям производства ($u=2$), v – объект одного из выбранных

срезом (учреждение или отрасль производства). Тогда, для Y , будет справедливо следующее (5-8):

$$u = 1 \Rightarrow v \in \{1, 2, \dots, 14\}, \quad (5)$$

$$u = 2 \Rightarrow v \in \{1, 2, \dots, 13\}, \quad (6)$$

$$u = 3 \Rightarrow v \in \{1, 2, \dots, 19\}, \quad (7)$$

$$u = 4 \Rightarrow v \in \{1, 2, \dots, 8\}. \quad (8)$$

Для срезов по учреждениям будут характерны группы $Y_{u \in \{1, 3, 4\}, v \in \{\dots\}}$, им будут соответствовать группы параметров $X_{m=1, p \in \{1, 2, \dots, 6\}, o \in \{\dots\}}$; $X_{m=2, p=1, o \in \{1, 2, \dots, 61\}}$; $X_{m=3, p=1, o \in \{1, 2, \dots, 155\}}$; $X_{m=4, p=1, o \in \{\dots\}}$. Срезы по отраслям производства можно охарактеризовать одной большой группой $Y_{u=2, v \in \{1, 2, \dots, 13\}}$, для которой будут соответствовать $X_{m=1, p=7, o \in \{1, 2, \dots, 24\}}$; $X_{m=2, p=2, o \in \{1, 2, \dots, 37\}}$. Структура индексов для Y представлена на рисунке 2.

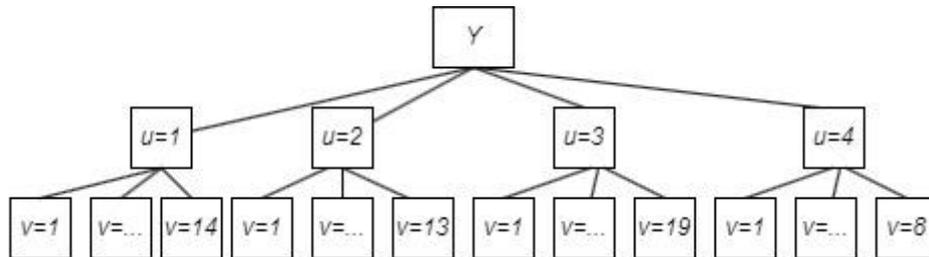


Рис. 2. – Структура индексов для возможных срезов параметров

Для времени t вместо индексации следует рассмотреть промежутки учета данных: за месяц, квартал или год. Учет данных за месяц: $t_{month \in \{1, 2, \dots, 12\}, year \in \{2014, \dots\}}$, квартал $t_{quarter \in \{1, 2, 3, 4\}, year \in \{2014, \dots\}}$, ГОД $t_{year \in \{2014, \dots\}}$. Структура индексации данных представлена на рисунке 3.

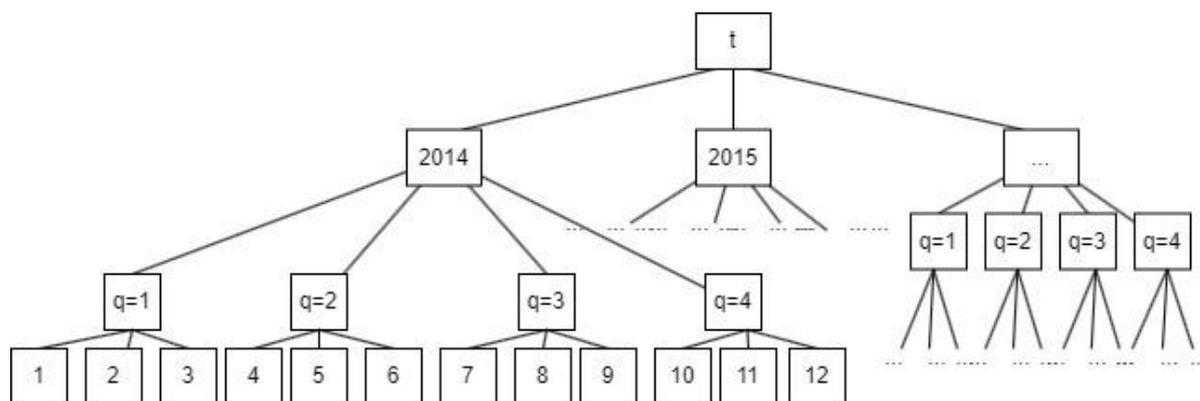


Рис. 3. – Структура индексов для периодов учета данных производственно-экономического и трудового секторов

Таким образом, у нас получена многомерная система индексации, благодаря которой возможно как записывать, так и извлекать нужные данные, рассматривать срезы, масштабировать информацию. Далее рассмотрим практический пример применения данной системы индексации.

Практический пример использования и *ORM*

Предположим, что нужно извлечь данные по среднесписочной численности работников, которые проходят трудовую адаптацию в августе 2021 года в исправительных колониях. Представленный запрос будет выглядеть следующим образом (9):

$$X_{m=1,p=1,o=1} \wedge Y_{u=1,v=4} \wedge t_{08.2021} \quad (9)$$

Если использовать пример извлечения данных на *Python*, то здесь в первую очередь следует руководствоваться правилами документа *PEP8*, т.е. названия индексов в виде переменных можно представить как: $X1_1_1$; $Y1_4$; $t8_2021$. Применяв язык запросов баз данных *SQL* [7, 8] и *ORM (Object-Relational Mapping)* [9, 10] библиотеки, такие как *SQLAlchemy*, данный запрос можно реализовать следующим образом (рисунок 4).

```
from sqlalchemy import create_engine, text

# Подключение к базе данных
engine = create_engine('строка подключения')

# SQL-запрос
sql_query = text("SELECT * FROM your_table WHERE X = :x AND Y = :y AND date_column = :date")
result = engine.execute(sql_query, x=X_1_1_1, y=Y_1_4, date=8_2021)
```

Рис. 4. – Пример использования системы индексации на *Python* с использованием библиотеки *SQLAlchemy*

В результате, представленную систему индексации данных возможно использовать не только на теоретическом уровне, но и с позиции практического применения. Использование *ORM*, а также запросов *SQL*, здесь может быть хорошим решением для дальнейшей практической реализации.

Заключение

Приведенные структуры и система индексации данных достаточно хорошо укладываются в парадигму объектно-ориентированного подхода. Используя приведенную систему индексов на языке *SQL*, можно оптимизировать извлечение и запись информации в базах данных. Также здесь особенно актуально использование *ORM* и сопутствующих библиотек.

Все это позволит создавать более структурированные базы данных, с возможностью более быстрой записи и чтения информации, а также более удобной работы с данными в целом. Актуальным продолжением представленной разработки может явиться проведение классового распределения исследуемых данных и создание *KPI (Key Performance Indicators)*, разработка *ER*-моделей и *OLAP*-кубов для практического применения на производствах.

Литература

1. Пономарев Д.С. Разработка алгоритмов обработки временных рядов при работе с статистическими отчетными формами производственного сектора пенитенциарной системы // Инженерный вестник Дона, 2024, №2. URL: ivdon.ru/ru/magazine/archive/n2y2024/9029
2. Царькова Е.Г. Методы поддержки принятия решений в деятельности органа исполнительной власти по обеспечению надежности ведомственной информационной среды // Инженерный вестник Дона, 2022, № 7. URL: ivdon.ru/ru/magazine/archive/n7y2022/7814.
3. Царькова Е.Г. Математическая модель управления надежностью системы интеллектуального видеонаблюдения на охраняемых объектах уголовно-исполнительной системы // Инженерный вестник Дона, 2022, № 8. URL: ivdon.ru/ru/magazine/archive/n8y2022/7854.
4. Пономарев С.Б., Горохов М. М., Серебренников А. В., Логинова С. Г. К вопросу о применении информационных систем для оптимизации тактики ведения больных в местах лишения свободы// Интеллектуальные системы в производстве, 2007. №2. С. 100–103.
5. Тененев В.А., Пономарев С.Б., Сергиенко А.С., Тоцкий С.И. Исследование критерия эффективности медицинского обслуживания для оптимизации сети медицинских частей учреждений территориального органа ФСИН России// Интеллектуальные системы в производстве, 2007. № 1(9). С.85-99.
6. Date C.J. SQL and Relational Theory. O'Reilly. 2010. 474 p
7. Кара-Ушанов В.Ю. SQL – язык реляционных баз данных. УрФУ, 2016.— 156 с.
8. Walter Shields SQL Quick Start Guide // Clyde Bank technology. 2022. 223 p.
9. Lutz M. Learning Python // 5th Edition Vol.1 O'Reilly, 2019. 832 p.



10. Cory Althoff. The Self-Taught Computer Scientist: The Beginner's Guide to Data Structures & Algorithms // Wiley, 1st edition. 2021. 224 p.

References

1. Ponomarev D.S. Inzhenernyj vestnik Dona, 2024, №2. URL: ivdon.ru/ru/magazine/archive/n2y2024/9029
2. Czar`kova E.G. Inzhenernyj vestnik Dona, 2022, № 7. URL: ivdon.ru/ru/magazine/archive/n7y2022/7814
3. Czar`kova E.G. Inzhenernyj vestnik Dona, 2022, № 8. URL: ivdon.ru/ru/magazine/archive/n8y2022/7854
4. Ponomarev S.B., Gorochov M. M., Serebrennikov A. V., Loginova S. G. Intellektual`ny`e sistemy` v proizvodstve, 2007. №2. pp. 100–103.
5. Tenenev V.A., Ponomarev S.B., Sergienko A.S., Toczkiy S.I. Intellektual`ny`e sistemy` v proizvodstve, 2007. № 1(9). pp. 85-99.
6. Date C.J. SQL and Relational Theory. O'Reilly. 2010. 474 p.
7. Kara-Ushanov V. Yu. SQL – yazyk`k relyacionny`x baz danny`x [SQL – relational database language]. UrFU, 2016. 156 p.
8. Walter Shields SQL Quick Start Guide. Clyde Bank technology. 2022. 223 p.
9. Lutz M. Learning Python. 5th Edition Vol.1 O'Reilly, 2019. 832 p.
10. Cory Althoff. The Self-Taught Computer Scientist: The Beginner's Guide to Data Structures & Algorithms. Wiley, 1st edition. 2021. 224 p.

Дата поступления: 27.01.2024

Дата публикации: 7.03.2024