

SQL³ КАК ИНСТРУМЕНТ ИНТЕГРАЦИИ И АНАЛИЗА ДАННЫХ В СОЦИАЛЬНОЙ АНАЛИТИКЕ

Исаченко А.А. студентка группы ИУ6-44Б

Московский государственный технический университет им. Н. Э. Баумана

*Научный руководитель: Шалдунова Т.Н., кандидат исторических наук,
доцент кафедры «Информационная аналитика и политические технологии»*

Аннотация: В статье рассматривается роль SQL в интеграции сведений из различных источников. Описываются шаги подготовки данных, создания схемы базы данных, загрузки и анализа информации с использованием Structured Query Language. Статья подчеркивает значение использования SQL для социального анализа.

Ключевые слова: данные, интеграция, анализ, SQL, база данных, социальная аналитика

Современный мир насыщен огромным объемом различных сведений, и социальная сфера не является исключением. Их сбор и анализ играют ключевую роль в принятии решений, направленных на улучшение качества жизни людей и оптимизацию социальных программ. Однако, эффективное управление информацией может быть сложной задачей из-за разнообразия её источников и форматов. Чтобы получить полное представление о происходящем, необходимо в совокупности рассматривать данные, собранные из множества ресурсов.

Интеграция представляет собой процесс объединения информации из различных источников в единый набор готовый к дальнейшему анализу. В социальной аналитике могут быть необходимы разные сведения, например, демографические и экономические показатели, их источниками может быть множество сфер, в частности, здравоохранения или образования. Эти данные могут поступать из опросов, исследовательских проектов, различных социальных сетей, таких как Facebook*, X*, ВКонтакте и других ресурсов. Для хранения и работы с большим количеством информации правильнее всего использовать одно централизованное хранилище.

База данных – это совокупность сведений, организованная в соответствии с определёнными правилами и поддерживаемая в памяти компьютера, характеризующая актуальное состояние некоторой предметной области. Существует множество способов работы с такими базами, например, можно использовать платформу для обработки и хранения информации Hadoop, различные языки программирования, среди кото-

³ SQL – Structured Query Language – язык программирования для работы с базами данных.

* С 2022 года запрещены в РФ согласно ч. 2 ст. 13.15 КоАП РФ

рых Python, R и Java, однако, в этой статье будет рассмотрена работа с языком программирования SQL. Он был создан сотрудниками компании IBM Дональдом Чемберлин и Раймондом Бойсом в 1970-х годах. SQL базы данных являются мощным инструментом для создания, изменения, управления и интеграции информации, полученной из различных ресурсов. Этот язык программирования позволяет создавать запросы, которые объединяют сведения разного формата из нескольких таблиц и источников в один набор. SQL предоставляет стандартизированный способ работы с данными, что упрощает процесс их интеграции и обработки.

Одним из главных преимуществ использования SQL в социальной аналитике является его гибкость и мощь при работе с большими объемами информации. Этот язык позволяет проводить различные операции, такие как объединение, фильтрация, агрегация и сортировка, что делает его идеальным инструментом для обработки сведений в социальной сфере. Кроме того, SQL базы данных обладают высокой производительностью и масштабируемостью – эти характеристики позволяют обрабатывать огромные объемы информации без потери скорости. Это особенно важно, когда данные очень обширные и детализированные. SQL также обеспечивает надежную защиту информации и управление доступом к ней, что важно для обработки конфиденциальных сведений, например, личных и связанных со здоровьем.

Шаги интеграции и анализа данных в SQL:

1. Создание схемы базы данных: для хранения информации в SQL необходимо создать соответствующую структуру базы. Этот процесс включает в себя определение таблиц, их полей (строк таблиц) и связей между ними. Например, для хранения информации о населении можно создать таблицы со сведениями о людях, их адресах, семейном статусе и других характеристиках, а также таблицы, содержащие данные о социальных программах и их участниках. Важно учесть необходимость нормализации базы данных для избежания избыточной информации и обеспечения её целостности.

2. Второй шаг – это подготовка к импорту в SQL базу данных. Этот этап включает в себя стандартизацию и преобразование информации в форматы, совместимые с используемым языком программирования. Например, сведения из различных источников могут иметь разные виды даты или кодировку символов, которые необходимо стандартизировать. Помимо этого, необходимо учитывать возможные проблемы совместимости и структурированности информации, так как они могут повлиять на последующий анализ.

3. Импорт и очистка: после создания схемы базы данных и проведения подготовительной работы можно приступить к загрузке сведе-

ний из различных источников. Это может быть реализовано с помощью различных SQL команд или специализированных инструментов для интеграции разнородной информации. Важно убедиться, что она загружается корректно и соответствует структуре базы данных. Также после импорта сведений их следует очистить от ошибок, дубликатов и пропущенных значений. SQL предоставляет множество инструментов для выполнения этой задачи.

4. Анализ: после загрузки данных в базу можно приступить к их оценке. Она включает в себя написание SQL запросов для извлечения нужной информации, проведение агрегации, фильтрации и детального рассмотрения сведений для выявления закономерностей и трендов. Например, можно исследовать социально-экономические показатели разных регионов и выявлять зависимости между ними. Помимо этого, SQL предоставляет возможность использования различных функций и операторов для проведения сложного анализа, включая группировку информации, вычисление статистических показателей и построение сводных таблиц.

5. Визуализация: после выполнения запросов данные можно визуализировать с помощью различных инструментов, таких как Tableau, Power BI, Excel и т.д. SQL также позволяет создавать отчеты и дашборды (информационные панели, которые получают сведения из разных систем и отображают их в понятном виде, на них используют текст, графики, диаграммы и др.) для удобного отображения и лучшей читаемости.

Базы данных SQL являются мощным инструментом для социального анализа. Правильное использование этого языка программирования позволяет объединять сведения из различных источников, проводить оценку и получать ценную информацию для улучшения качества жизни людей и общества в целом. Для достижения оптимальных результатов важно помнить о необходимости некоторой предварительной работы, включающей в себя подготовку данных и создание их правильной структуры в базе. При этом необходимо учитывать разнообразие информационных ресурсов, особенности их структуры и формата, а также проводить адекватную интерпретацию результатов анализа с учетом контекста и специфики социальной сферы.

Литература и источники

1. Бочарников И.В. Информационно-аналитические технологии противодействия дестабилизации внутривнутриполитической ситуации в России. В сборнике: Прогнозируемые вызовы и угрозы национальной безопасности Российской Федерации и направления их нейтрализации. Сборник материалов круглого стола. Москва, 2021. С. 135–157.

2. Девять языков для работы с Big Data. <https://blog.skillfactory.ru/9-yazykov-dlya-raboty-s-big-data/>.

3. Карпов, Д., Уайт, С. Обработка баз данных на SQL. Москва: Питер, 2018. 352 с.
4. Ремарчук В.Н. Социально-гуманитарное образование в технических вузах на пути перемен. В сборнике: Интеграция образования, науки и производства в интересах высокотехнологического комплекса. Материалы Международного форума «Технологии в машиностроении – 2010». 2010. С. 62–67.
5. Фомин М.М. Реляционные базы данных. Учебное пособие для бакалавров. Москва: МГТУ им. Н.Э. Баумана, 2023. 161 с.
6. Шалдунова Т.Н. Политика безопасности: Россия в современном мире // Вопросы национальных и федеративных отношений. 2022. Т. 12. № 5 (86). С. 1755–1760.
7. Шилдс У. SQL: быстрое погружение. Москва: Издательский дом Вильямс, 2005. 416 с.
8. Content of the process of formation of students' speech abilities at the university / Ovsyannikova O.A., Mishcherina M.A., Vocharnikov I.V. В сборнике: E3S Web of Conferences. 8. Сер. "Innovative Technologies in Science and Education, ITSE 2020" 2020. С. 18106.