

Метаданные (Metadata)

Синонимы: Данные о данных

Метаданные — это средство классификации, упорядочивания и характеристики данных, или, говоря простыми словами, — «данные о данных». Они несут информацию о типе, структуре, происхождении, свойствах, авторе, владельце, месте и времени создания и другие сведения, которые необходимы для управления основными и справочными данными компании. Метаданные явно или неявно присутствуют во всех информационных объектах, имеющих определенную структуру — файлах документов, изображений, веб-страниц, видео и аудио материалов, баз и хранилищ данных и др.

Метаданные являются важнейшим элементом структуры корпоративных данных. Они позволяют оценить актуальность той или иной информации, определить цели и способы ее обработки, обеспечивают интерфейс пользователей аналитических платформ и приложений, образуя семантический слой, который дает возможность оперировать бизнес-терминами при формировании сложных аналитических запросов.

Выделяют несколько видов метаданных:

1. **Описательные** — содержат общую информацию об объекте, такую как время и место его создания, идентификатор, заголовок, размер, сведения о владельце и т.д.
2. **Структурные** — описывают структуру объекта. Для текстового документа это может быть разбиение на страницы, абзацы, разделы, главы. Для таблицы — имена и типы столбцов, допустимое число символов в каждом.
3. **Административные** — необходимы для управления информационными ресурсами компании. Они включают данные о правах доступа, политике безопасности, регламенте обработки и т.д.
4. **Технические** — используются для управления процессами сбора, хранения, обработки, анализа, представления и обеспечения безопасности корпоративных материалов.
5. **Справочные** — содержат информацию об объектах данных, например коды документов.
6. **Статистические или процессные** — содержат статистику, связанную с процессом сбора, обработки и хранения корпоративной информации.
7. **Юридические** — описывают сведения о создателе, правообладателе, лицензировании и других юридических аспектах использования материалов.

Важнейшую роль метаданные играют в организации поиска контента в Интернете. Каждая веб-страница включает информацию, определяющую язык и инструмент, с помощью которых она создана, темы, которым она посвящена, и с какими другими

страницами связана.

Цифровое изображение может включать метаданные, описывающие его размер, глубину цвета, разрешение, выдержку и другие сведения.

Метаданные текстового файла могут содержать информацию о числе структурных единиц документа (символов, предложений, абзацев, страниц, глав и т.д.), об авторах, времени создания, аннотацию и ключевые слова.

Для эффективного управления метаданными в корпоративных информационных системах создаются специальные реестры и репозитории. Вторые отвечают за физическое хранение и каталогизацию. Данные в них должны быть общедоступными, интегрированными, актуальными и историческими. Репозиторий позволяет объединить метаданные всех областей бизнеса и поддерживает управление информацией в масштабе предприятия, ее качество и интеграцию в системах поддержки принятия решений.

В отличие от общедоступного репозитория, реестр метаданных является защищенной структурой, изменения в которую могут вносить только уполномоченные лица. Он хранит не только сами метаданные, но и связи между ними.

Таким образом, репозитории метаданных ориентированы на поддержку бизнес-аналитики и принятие решений, а реестр — на управление корпоративными данными (сбор, хранение, интегрирование и т.д.).