

Тема 3. Метадані. Класифікація метаданих

План

1. Поняття метаданих.
2. Класифікація метаданих.
3. Формат метаданих.

Мета вивчення теми: засвоїти поняття «метадані» та особливості роботи з ними.

Перелік ключових слів та понять із теми

Метадані, метадані за змістом, метадані до ресурсу, формат метаданих

Теоретичні відомості з теми

1. Поняття метаданих

Метадані (у загальному випадку) – це дані, що характеризують або пояснюють інші дані. Наприклад, значення «123456» само по собі недостатньо виразне. А якщо значенню «123456» зіставлено ім'я «поштовий індекс» (що вже є метаданими), то в цьому контексті значення «123456» більш осмислене – можна «витягувати» інформацію про місцеположення адресата, що має даний поштовий індекс.

Оскільки для більшості людей різниця між словами «дані» та «інформація» існує тільки з філософської точки зору і не істотна з практичної точки зору, то мають місце такі визначення:

- метадані – це інформація про дані;
- метадані – це інформація про інформацію.

Для терміну **метадані** немає єдиного формального визначення. Навпаки, існують різні визначення цього терміну. Ось просте і популярне переформулювання:

Метадані – це дані про дані. Цей термін в широкому сенсі слова використовується для будь-яких «даних про дані»: іменах таблиць, колонок в таблиці, програм і тому подібне.

Метадані – це дані з більш загальної формальної системи, що описує задану систему даних.

Існують **вужчі визначення**:

Метадані – це структуровані дані, що представляють собою характеристики описуваних сутностей для цілей їх ідентифікації, пошуку, оцінки, управління ними.

Метадані – це набір допустимих структурованих описів, які доступні в явному вигляді і призначення яких допомогти знайти об'єкт. Це визначення використовується набагато рідше, оскільки воно концентрується на одному з

призначень метаданих – пошук об'єктів, сутностей, ресурсів – та ігнорує інші призначення.

Відмінність між даними і метаданими. Зазвичай неможливо провести однозначне розділення на дані та **метадані** в документі, оскільки:

1. Щось може бути як даними, так і метаданими. Так, заголовок статті можна одночасно віднести як до метаданих (як елемент метаданих – заголовок), так і до власне даних (оскільки заголовок є частиною самого тексту).

2. Дані та метадані можуть мінятися ролями. На вірш, що розглядається як дані, може бути написана музика, в цьому випадку весь вірш може бути «прикріплений» до музичного файлу і в цьому випадку розглядається як метадані. Отже, віднесення до однієї або іншої категорії залежить від точки зору.

3. Можливе створення мета-мета-...-метаданих. Оскільки, відповідно до звичайного визначення, метадані є даними, то можна створити метадані на метадані, метадані на метадані на метадані і т.д. На перший погляд це може здатися безглуздим, але насправді це є дуже істотною і корисною властивістю даних і метаданих.

Ці міркування застосовні незалежно від вибору визначення метаданих (з приведених вище і не тільки).

Метадані використовуються для підвищення якості пошуку. Пошукові запити, використовуючи метадані, можуть врятувати користувача від зайвої ручної роботи з фільтрації. Інформуючи комп'ютер про те, які елементи даних зв'язані і як ці зв'язки враховувати, стає можливим здійснювати достатньо складні операції з фільтрації та пошуку. Наприклад, якщо пошукова система «знає» про те, що «Ван Гог» є «голландським художником», то вона може видати у відповідь на запит про голландських художників веб-сторінку про Ван Гога, навіть якщо слова «голландський художник» не зустрічаються на цій сторінці.

Практично кожний електронний документ має певні метадані. Метадані електронних документів відіграють важливу роль у системах електронного документообігу та автоматизації діловодства, в інформаційно-пошукових системах. Метадані можуть, наприклад, включати дату, коли документ був збережений і відомості про особистість користувача, що зберіг його. Системи електронного документообігу та автоматизації діловодства можуть також здобувати метадані з документу автоматично або підказувати користувачеві додати метадані.

2. Класифікація метаданих

Метадані можна класифікувати за такими ознаками:

- *Змістом.* Метадані можуть або описувати сам ресурс (наприклад, назва і розмір файлу), або вміст ресурсу (наприклад, «у цьому відеофайлі показано, як хлопець грає у футбол»).

- *Відношенням до ресурсу в цілому.* Метадані можуть відноситися до ресурсу в цілому або до його частин. Наприклад, «Title» (назва фільму) відноситься до фільму в цілому, а «Scene description» (опис епізоду фільму) окреме для кожного епізоду фільму.

- *Можливістю логічного виводу.* Метадані можна підрозділити на три шари: нижній шар – це «сирі» дані самі по собі; середній шар – метадані, що описують ці дані; і верхній шар – метадані, які дозволяють робити логічний вивід, використовуючи другий шар.

3. Формат метаданих

Метаданими на практиці зазвичай називають дані, представлені відповідно до одного з форматів метаданих.

Формат метаданих – це стандарт, призначений для формального опису деякої категорії ресурсів (об'єктів, сутностей, документів і т.п.). Такий стандарт зазвичай включає набір полів (атрибутів, властивостей, елементів метаданих), що дозволяють характеризувати даний об'єкт. Наприклад, формат MARC дозволяє описувати книги (і не тільки книги), містить поля для опису назви, автора, тематики і безлічі інших характеристик (формат MARC дозволяє описати сотні характеристик).

Формати метаданих часто розробляються міжнародними організаціями або консорціумами, що включають зацікавлені в появі стандарту державні організації та приватні компанії. Розроблений формат часто закріплюється як стандарт в одній або декількох організаціях, що займаються розробкою й ухваленням стандартів (наприклад W3C, ISO, ANSI тощо).

Питання для самоконтролю

1. Дайте визначення поняттю метадані.
2. Які ви знаєте відмінності між даними і метаданими?
3. Для чого використовуються метадані?
4. Наведіть класифікацію метаданих.
5. Дайте визначення формату метаданих.