

Проблеми та визначення якості консолідованих даних

Думанський Нестор
Кафедра СКІД
НУ “Львівська політехніка”
Львів, Україна
nestor.o.dumanskyi@lpnu.ua

Видойник Марія
Кафедра СКІД
НУ “Львівська політехніка”
Львів, Україна
maria.vidoynik@gmail.com

The analysis of the problems that arise in the process of consolidation of information. The paper formalizes quality characteristics of consolidated data in dataspace. There is told the notion of usefulness of data from data sources. There is analysed an architecture for a system of quality evaluation of different types of data.

Ключові слова: консолідована інформація, консолідація, консолідовані дані, інформаційні продукти, якість даних, корисність даних, характеристики якості.

ВСТУП

Сьогодні очевидною стала перевага інформаційної складової діяльності людей над всіма іншими її формами і компонентами. Д. Прайс ще наприкінці 50-х років минулого століття обґрунтував модель прискореного росту документального потоку. Згідно з нею кількість документів подвоюється кожні 10-15 років [1].

Такий стан справ спричиняє об'єднання даних, розміщених у різних, наперед неузгоджених джерелах, що має назву *консолідація*.

Консолідовані дані – це одержані з декількох джерел та системно інтегровані різнотипні інформаційні ресурси, які в сукупності наділені ознаками повноти, цілісності, несуперечності та складають адекватну інформаційну модель проблемної області з метою її аналізу, опрацювання та ефективного використання в процесах підтримки прийняття рішень [2].

Такий процес потребує застосування різноманітних технологій пошуку, накопичення, опрацювання та зберігання інформації, внаслідок чого актуальним стає питання якості даних.

ОСНОВНА ЧАСТИНА

Створення таких інформаційних продуктів відбувається у декілька етапів [3]:

- визначення цілей і планування з зазначенням необхідних термінів;
- пошук та збір даних;
- попереднє опрацювання, структурування даних, тобто перетворення на інформацію;
- аналіз і синтез інформації – перетворення на знання;
- формування інформаційних продуктів, які будуть задовольняти потенційних споживачів.

Незалежно від сфери застосування консолідація інформаційних ресурсів позначає об'єднання декількох елементів у єдиний об'єкт, щоб досягнути певної загальної мети.

Для подальшого ефективного використання та прийняття рішень у будь-якій галузі необхідно, щоб дані відповідали таким вимогам:

- повнота, несуперечливість та вчасне надходження;
- однотипна структура, що забезпечить можливість завантаження у єдине сховище даних та аналіз;
- інформативність, адже на їх основі відбуватиметься прийняття рішень;
- зберігання в однакових моделях даних;
- незалежність від платформи розроблення для реалізації можливості їх використання іншими засобами.

Під *якістю даних* мається на увазі процес компонування даних, що включає імена ресурсів, схеми даних, особливості доступу і опрацювання, очищення, вдосконалення даних, а також

об'єднання з усуненням дублювання та певної невизначеності.

Корисність даних для користувачів залежить від ступеня довіри до джерела даних:

$$Trust_j = \frac{\sum_{i=1}^m (Trust_k(i, j))}{(n * m)}, \quad (1)$$

де n – кількість звернень користувача до ресурсу; m – кількість користувачів, що звертались до ресурсу; $Trust_k(i, j)$ – значення лінгвістичної змінної, що відображає довіру i -го користувача до j -го джерела даних при k -ому зверненні.

Базовими показниками якості даних згідно зі стандартом ISO 9126 є: функціональна придатність до використання; коректність або достовірність; ресурсна економічність; практичність; супроводжуваність; мобільність.

Функціональна придатність – відношення повноти накопичених об'єктів у джерелах даних, до загальної кількості об'єктів, що потрапили у локальне сховище.

Коректність або *достовірність даних* – це ступінь відповідності даних про об'єкти в базах даних реальним об'єктам в даний момент часу.

Ресурсна економічність відображає зайнятість ресурсів центрального процесора, оперативної, зовнішньої та віртуальної пам'яті, каналів введення-виведення, терміналів і каналів зв'язку.

Практичність – функціональна придатність і корисність застосування консолідованих даних для певних користувачів.

Супроводжуваність визначається зручністю і ефективністю виправлення, удосконалення або адаптації структури та змісту описаних даних залежно від змін у зовнішньому середовищі застосування, а також у вимогах і функціональних специфікаціях замовника.

Мобільність характеризується тривалістю і трудомісткістю інсталяції даних, адаптації та заміщення при перенесенні на інші апаратні та операційні платформи.

Для визначення ресурсної економічності, практичності, супроводжуваності існують розроблені методи та засоби.

Таким чином, *якість консолідованих даних* обчислюється як інтегральна характеристика, що відображає повноту накопичення даних, коректність, мобільність та корисність прийнятих рішень:

$$s_1 \text{quality} + s_2 v(r) \rightarrow \text{Max}, \quad (2)$$

де $quality$ – інтегральний безрозмірний показник характеристик якості даних, $0 \leq \text{quality} \leq 1$, s_1 – коефіцієнт важливості повноти накопичення даних; $v(r)$ – значення багатомірної функції корисності; s_2 – коефіцієнт важливості якості прийнятих рішень, $s_1 + s_2 = 1$.

ВИСНОВОК

Коли дані використовуються з декількох джерел, що є неузгодженими наперед, їх якість різко знижується і вже не може задовольняти потреби користувачів через неузгодженість структури, показників, форматів, різне представлення тощо. Компонування даних, очищення та вдосконалення, а також об'єднання з усуненням дублювання та невизначеності становить процес оцінювання якості даних.

Підсумовуючи, варто наголосити, що інформаційні технології неперервно розвиваються, набувають глобалізаційного характеру. Саме цінною є інформація, яка отримана з різних джерел, проте наділена ознаками цілісності та несуперечливості.

ЛІТЕРАТУРА

- [1] Матвієнко О.В. Консолідована інформація: навч. посіб. / О.В. Матвієнко, М.Н. Цивін. – К.: “Центр учбової літератури”, 2014. – 134 с.
- [2] Шаховська Н.Б. Проблеми якості консолідованих даних у просторах даних / Н.Б. Шаховська, О.Ю. Пшеничний, І.М. Чорней. – Львів : Львівська політехніка, 2011. – 80 с.
- [3] Кунанець Н.Е., Пасічник В.В. Консолідація інформації та інформаційна безпека // III міжнародна НПК “Інформаційна та економічна безпека” : тез. доп. – Львів: Львівська політехніка, 2010. – 131.