

ДОСЛІДЖЕННЯ ЯКОСТІ КОНСОЛІДОВАНИХ ДАНИХ У ПРОСТОРАХ ДАНИХ

Анотація. У статті формалізовано характеристики якості консолідованих даних у просторах даних. Введено поняття корисності даних з джерел даних. Введено концептуальне визначення якості консолідованих даних простору даних. Розроблено метамови опису джерел даних та встановлення відповідності між їхніми структурами даних. Розроблено архітектури підсистеми оцінки якості консолідованих даних.

Ключові слова: простір даних, консолідація, якість даних.

Аннотация. В статье формализованы характеристики качества консолидированных данных в пространствах данных. Введено понятие полезности данных с источников данных. Введено концептуальное определение качества консолидированных данных пространства данных. Разработан метаязык описания источников данных и установления соответствия между их структурами данных. Разработана архитектура подсистемы оценки качества консолидированных данных.

Ключевые слова: пространство данных, консолидация, качество данных.

Abstract. The quality characteristics of consolidated data in data spaces are formalized in the article. A concept utility data from data sources was introduced. A conceptual definition of quality consolidated data in data space was introduced. Metalanguages of data sources were developed and matches between their data structures were given. Architecture subsystems quality assessment of consolidated data was introduced.

Keywords: dataspace, consolidation, data quality.

1. Вступ

Інформаційне суспільство – суспільство, в якому створення, поширення, дифузія, використання, інтеграція і маніпулювання інформації – важлива господарська, політична і культурна діяльність.

Специфікою цього виду суспільства є те, що задача консолідації даних (об'єднання даних, розміщених у різних, наперед не узгоджених джерелах) виникає досить часто. Так, для університету прикладом консолідації є формування наукових звітів, визначення показників успішності та якості навчання, формування рейтингу кафедри тощо; для туристичної сфери – це визначення критичних показників розвитку регіону.

Для прийняття адекватних рішень у певній галузі необхідно, щоб дані, які надаються різними інформаційними ресурсами та у подальшому використовуються для прийняття керівних рішень, задовольняли таким вимогам:

- були повними, несуперечливими та надходили вчасно;
- були інформативними, оскільки вони застосовуватимуться для прийняття рішень;
- були однакової структури, щоб мати можливість завантажити їх у єдине сховище даних та проаналізувати;
- зберігалися в однакових моделях даних та були незалежними від платформи розроблення, щоб мати можливість використання цих даних іншими засобами.

Тому при опрацюванні даних з різних джерел виникає проблема якості цих даних (відповідності даних вимогам користувачів). На рівні задач, для яких використовується точкове джерело, якість даних цього джерела є достатньою і задовольняє (повністю чи частково) потреби осіб, що приймають рішення на їх основі. Проте, коли йде мова про використання даних з декількох джерел, наперед не узгоджених, якість таких даних різко знижу-

ється і вже не може задовольняти потреб користувача через неузгодженість форматів, різне представлення тощо.

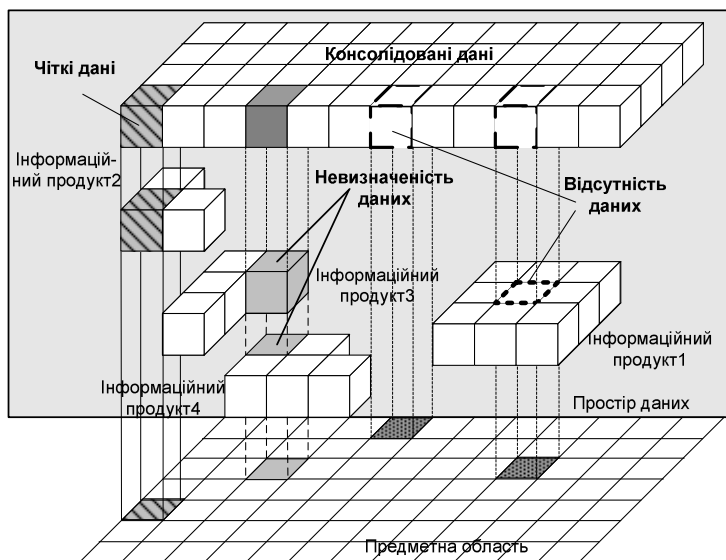
2. Аналіз літературних джерел

Над опрацюванням різнотипних даних працювали Colin White, A. Sheth, J. Larson, Maurizio Lenzerini, Frederick Lane, Christoph Koch, Xin Dong, Л.А. Калиниченко, С.А. Ступников, А.В. Фомичев, М.Н. Гриньов, С.Д. Кузнецов та ін. Розроблені моделі та метамови опрацювання різнотипних даних [1–9]. Проте вказані моделі та методи опрацьовують або лише наперед відомі типи даних (здебільшого, реляційні бази даних), або вирішують лише часткові проблеми опрацювання різнотипних даних, наприклад, індексування для пришвидшення пошуку. Тому на сьогодні немає жодної методики опрацювання даних, яка б задовольняла всім наведеним вимогам до якості даних, а отже, немає можливості аналізувати стан галузі загалом, використовуючи першоджерела інформації, а не визначені наперед статистичні звіти.

Отже, в статті необхідно розв'язати науково-практичну проблему підвищення якості консолідованих даних з різнотипних, неузгоджених інформаційних продуктів певної предметної області з метою подальшого прийняття керівних рішень на їх основі.

3. Постановка задачі

Інформаційні продукти, що описують певну предметну область, та консолідовані дані становитимуть простір даних. Однією із задач, яка виникатиме у процесі консолідації, є невизначеність даних, що є результатом дублювання, неточності, відсутності, протиріччя даних (рис. 1).



розроблення словника даних дозволить уникнути цієї проблеми та частково уніфікувати схеми джерел даних.

Під якістю даних розумітимемо процес компонування даних (включаючи імена ресурсів, схеми даних, особливості доступу, опрацювання тощо), очищення та вдосконалення даних, а також об'єднання з усуненням дублювання й невизначеності.

До якісних даних ставляться такі вимоги: повнота, точність, зв'язність, доступність, актуальність, своєчасність.

Наступною задачею є визначення залежності якості стратегічних рішень від кількості джерел, на основі яких рішення приймаються. Але, оскільки для прийняття рішення

Іншою задачею є визначення та узгодження схем даних інформаційних ресурсів. Існуючі на сьогодні методи опрацьовують або наперед відомі схеми даних, або вимагають, щоб джерела даних (інформаційні продукти) перебували під жорстким контролем, що не дає змоги встановлювати змінні семантичні зв'язки. Також однією з перепон використання проаналізованих методів інтеграції є те, що розробники не завжди дотримувалися стандартів під час розроблення схем даних. Аналіз можливостей застосування існуючих стандартів показав, що

важливими є не тільки кількісні характеристики, але й якісні (корисність, довіра до джерела тощо), то необхідно також визначати ступінь довіри до джерела даних.

4. Формалізація корисності даних

Консолідація даних здійснюється на основі аналізу даних інформаційних продуктів, що входять у простір даних.

4.1. Визначення корисності даних з джерела для якості прийняття рішення

Оскільки основою побудови простору даних є підтримка подальшого процесу прийняття рішень на основі консолідованих даних, то необхідно проаналізувати вплив цих даних на якість прийнятого рішення. Критерій кращого чи гіршого рішення залежить від предметної області та конкретної задачі. Прикладами критеріїв є: співпадіння прогнозованого плану з реальним, мінімізація кількості вхідних даних, згортка параметрів тощо.

Корисність даних для певного користувача чи групи користувачів залежить також і від ступеня довіри до джерела даних.

Тоді визначення ступеня довіри i -го користувача до j -го джерела даних здійснюється за такою формулою:

$$Trust(i, j) = \frac{\sum_{k=1}^n Trust_k(i, j)}{n}, \quad (1)$$

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

Для розрахунку загального ступеня довіри до джерела j узагальнимо формулу (1):

$$Trust_j = \frac{\sum_{i=1}^m \left(\frac{\sum_{k=1}^n Trust_k(i, j)}{n} \right)}{n * m}, \quad (2)$$

де m – кількість користувачів, що звертались до ресурсу.

Ступінь довіри може встановлюватись і до конкретної характеристики джерела даних. Тоді він враховуватиме і ступінь довіри до джерела загалом, і довіру до конкретної характеристики:

$$Trust^{attr}(i, j) = Trust(i, j) * \frac{\sum_{k=1}^n Trust_k^{attr}(i, j)}{n}, \quad (3)$$

де $attr$ – назва атрибута, для якого здійснюється визначення ступеня довіри.

Визначимо корисність даних для прийняття рішення.

Нехай є критерій $R_j \in R$ оцінки наслідків рішення $x = (x_1, \dots, x_j, \dots, x_n)$, розподіл значень якого залежить тільки від компоненти x_j альтернативи x .

Якщо має місце незалежність критеріїв R_1, R_2, \dots, R_m за перевагою, то багатовимірна функція корисності прийнятого рішення $v(r)$ представлена у вигляді

$$v(r) = \sum_{j=1}^m k_j v_j(r_j). \quad (4)$$

Функцію v_j , що виражає оцінку значення r_j , можна вважати j -ою компонентою функції корисності, а k_j – вагою, що визначає критерій R_j . У випадку просторів даних вага джерела даних j визначається як $k_j = Trust(i, j)$ (див. (1)), де i є заданим і вказує на конкретного користувача.

4.2. Концептуальне визначення якості консолідованих даних простору даних

У просторах даних домінуючого значення набувають самі дані, їхнє зберігання і опрацювання. Для оцінювання якості даних застосуємо загальний методичний підхід до виділення адекватної номенклатури стандартизованих в ISO 9126 базових характеристик і субхарактеристик.

Базовими характеристиками якості даних є функціональна придатність до використання, коректність або достовірність, ресурсна економічність, практичність, супроводжувальність, мобільність.

Функціональна придатність даних просторів даних може подавати складну проблему для виміру та оцінки відповідності вимогам реальних значень атрибутів якості. Особливо це актуально для розподілених систем, у яких циркулює різноманітна і складна інформація про аналізовані об'єкти. Оскільки методи інтеграції, які застосовуються до сховищ даних, не можуть застосовуватись до просторів даних, то визначення функціональної придатності є однією з базових характеристик, що досліджується у роботі.

Функціональна придатність визначається, у першу чергу, повнотою накопичених об'єктів – відносною кількістю об'єктів або документів, наявних у джерелах даних, до загальної кількості об'єктів, що потрапили у локальне сховище, яке містить консолідовані дані:

$$plenitude = \frac{Count(local_dataware)}{\sum_i Count(source_i)}. \quad (5)$$

Коректність або достовірність даних – це ступінь відповідності даних про об'єкти в базах даних реальним об'єктам у даний момент часу, що визначається змінами самих об'єктів, некоректних записів про їх стан або некоректними розрахунками їх характеристик. Вибір та встановлення вимог до коректності даних можна оцінювати за ступенем покриття накопиченими, актуальними і достовірними даними стану і зміни зовнішніх об'єктів, які вони відображають. Оскільки у роботі джерелами даних є не об'єкти предметної області, а інформаційні продукти, то під коректністю даних будемо розуміти кількісну характеристику, що відображає відносну кількість описів об'єктів з джерел даних, які не містять дефектів і помилок, до загальної кількості об'єктів у просторі даних. У роботі коректність визначено як відносну кількість описів об'єктів, які не містять дефектів і помилок, до загальної кількості об'єктів у просторі даних:

$$identity = \frac{Count(\sigma_{Trust>0.6}(local_dataware))}{Count(local_dataware)}. \quad (6)$$

Використовуваність ресурсів (або ресурсна економічність) у стандартах відображається зайнятістю ресурсів центрального процесора, оперативної, зовнішньої та віртуальної пам'яті, каналів введення-виведення, терміналів і каналів зв'язку. Цей показник у роботі не буде аналізуватися, оскільки існують розроблені методи (наприклад, метод критичних робіт [10]) та засоби визначення завантаженості ресурсів.

Практичність – важко формалізоване поняття, яке визначає функціональну придатність і корисність застосування консолідованих даних для певних користувачів. Деякі субхарактеристики можна оцінювати економічними показниками – витратами праці і часу спеціалістів на реалізацію певних функцій взаємодії з даними. Оцінки практичності залежать не тільки від власних характеристик простору даних, але також від організації та адекватності документування процесів їх експлуатації. У просторі даних оцінка практичності здійснюватиметься за допомогою функції корисності прийнятих рішень (4).

Супроводжуваність даних відображається зручністю і ефективністю виправлення, удосконалення або адаптації структури та змісту описів даних залежно від змін у зовнішньому середовищі застосування, а також у вимогах і функціональних специфікаціях замовника. Для оцінки супроводжуваності розроблені методи та засоби (наприклад, технологія ETL – витягнення, трансформування, завантаження), тому у роботі ця характеристика даних не розглядається.

Мобільність характеризується тривалістю і трудомісткістю їх інсталяції, адаптації та заміщення при перенесенні на інші апаратні та операційні платформи. Інформація про процеси, що відбуваються у зовнішньому середовищі, може мати великі обсяги і трудомісткість первинного накопичення та актуалізації, що визначає необхідність її ретельного зберігання й регламентованих змін. Можливі ситуації, коли подібні дані є унікальними і невідновлюваними. У просторах даних характеристика мобільності пов'язана зі зміною даних про джерела даних у каталозі:

$$actuality = \frac{Count(\sigma_{metadata.time_update < datadiff("d", Now(), 30)}(local_dataware))}{Count(local_dataware)}. \quad (7)$$

Отже, під якістю консолідованих даних у просторі даних будемо розуміти інтегральну характеристику, яка відображає повноту накопичення даних (результат застосування методів консолідації), коректність, мобільність та корисність прийнятих рішень:

$$s_1 quality + s_2 v(r) \rightarrow Max, \quad (8)$$

де *quality* – інтегральний безрозмірний показник характеристик якості даних;

$$0 \leq quality \leq 1;$$

$$quality = n1 * plenitude + n2 * identy + n3 * actuality ;$$

s_1 – коефіцієнт важливості повноти накопичення даних;

$v(r)$ – значення багатовимірної функції корисності;

s_2 – коефіцієнт важливості якості прийнятих рішень, $s_1 + s_2 = 1$.

У наступних розділах будуть показані числові значення коефіцієнтів залежно від предметних областей, для яких будується простір даних.

4.3. Розроблення метамови опису джерел даних та встановлення відповідності між їхніми структурами даних

Для інтеграції даних у сховищі консолідованих даних *Cg* використовується каталог, схема якого подана на рис. 2.

Опишемо елементи метамови простору даних. Вважатимемо, що запит *q* до простору даних заданий коректно, якщо він складається з елементів, описаних у *Cg* та *Dic*.

$$q_{object(c_1, \dots, c_n)} : par = \left\{ \begin{array}{l} object \in Cg, Trust_{object} > 0, \\ (c_1, \dots, c_n) \in Dic \end{array} \right\} : par ,$$

де *object* – об’єкт, про який йде мова у запиті, (c_1, \dots, c_n) – назви характеристик об’єкта, *par* – список параметрів запиту. Залежно від типу джерела параметри можуть відігравати роль: параметрів пошуку – у текстових даних; умови вибору – для структурованих даних.



Рис. 2. Схема каталогу простору даних

Код метаданих дорівнює коду джерела, описаного у словнику термінів *Dis* на рис. 3.

Приклад: Нехай є такий запит – знайти усі старовинні замки Львівської області. Тоді на метамові він запишеться:

*q*_{старовинний (замок)} : Львівська область .

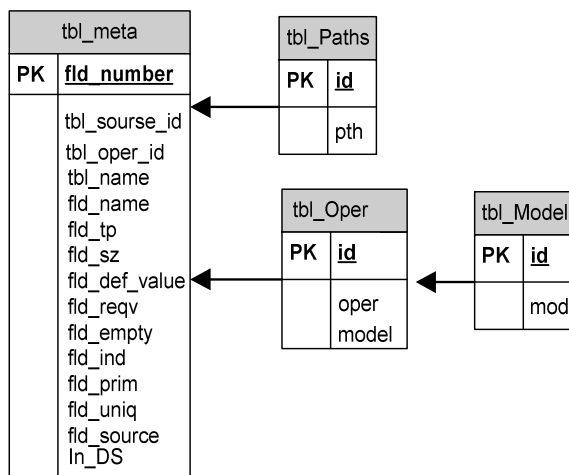


Рис. 3. Схема відношень бази знань для аналізу структури джерел даних

Трансформація метамови у запити до каталогу даних та синонімічного словника відбувається таким чином. Вибираємо з каталогу даних відомості про всі джерела, призначення яких – «старовинний» (рис. 3) та у яких Код = Код пов'язаного джерела. Якщо запит повернув результат, розраховуємо ступінь довіри до отриманих джерел і вибираємо ті, де ступінь довіри більший за нуль. Далі формуємо запит до словника даних і шукаємо синоніми до характеристики «замок».

Здійснюється це таким чином:

1. Пошук синонімічної назви «замок».
2. Вибір усіх синонімів, де значення «Посилання» таке ж, як і в синонімічній назви «замок».

Нехай маємо таку вибірку зі словника синонімів (рис. 4).

Синонімічна назва	Призначення	Аналог	Назва стандарту	Посилання
Замок	Старовинний	Castle		Castle1
Назва замку	Старовинний	Castle_name		Castle1
Назва	Старовинний	Name		Castle1
Об’єкт	Старовинний	Object	Dublin Core	Castle1
Код	Старовинний	Id	Dublin Core	Castle1

Рис. 4. Результат пошуку у синонімічному словнику

У результаті формуються запити до відібраних джерел за характеристиками, вказаними у вибірці зі словника. Вибірка буде формуватися на основі схеми даних, поданої на рис. 3.

Як видно з прикладу, алфавіт запиту об'єднує алфавіт усіх джерел даних, до яких направляють запит, а для встановлення характеристик вибираються усі можливі синоніми:

$$Dic = \{R\} \cup \{Rel\} \cup \{key\} \cup \{H\}, \quad (9)$$

де описані схеми баз даних, сховищ даних, ключових слів текстових файлів, заголовків веб-документів відповідно.

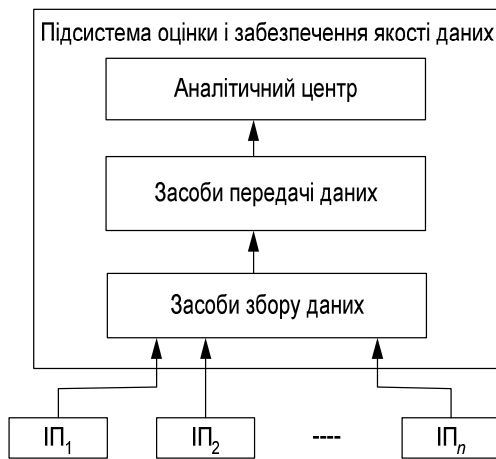


Рис. 5. Архітектура підсистеми оцінки і забезпечення якості консолідованих даних

4.4. Розроблення архітектури підсистеми оцінки якості консолідованих даних

Розподілена підсистема забезпечення і підтримки якості даних у просторі даних призначена для реалізації алгоритмів і процедур, які забезпечують оцінку якості даних, збір та обробку інформації для підтримки якості даних. Відповідно з цим визначенням до складу підсистеми входять аналітичний центр, засоби збору та передачі даних (рис. 5).

Основні завдання, які вирішує підсистема:

- проведення розрахунків з оцінки якості консолідованих даних;
- обчислення параметрів процедур для підтримки якості консолідованих даних у необхідному стані;

– збір, передача і обробка необхідної інформації для оцінки та забезпечення якості консолідованих даних;

– зберігання модифікацій інформаційних продуктів у каталозі простору даних для забезпечення якості (введення нових даних, видалення невідповідних даних).

Зупинимось на перелічених завданнях докладніше.

Розрахунки з оцінки якості консолідованих даних проводяться з урахуванням необхідних показників якості (наприклад, повнота та актуальність даних). У результаті проведення розрахунків визначаються значення кількісних показників якості.

Розрахунки для оцінки якості консолідованих даних повинні проводитися при вне-

сенні змін до складу даних (додавання або вилучення) або через певні інтервали часу, величина яких залежить від динаміки зміни кількості об'єктів. Визначення інтервалу часу відбувається на основі динамічної вибірки з відношення мета-даних джерела (рис. 4).

При оцінці якості консолідованих даних необхідно оцінювати якість даних, отриманих із джерел даних, та релевантність знайдених за запитом користувача. Якість даних, отриманих з джерел даних, не може бути покращено, оскільки формаційні продукти, що входять

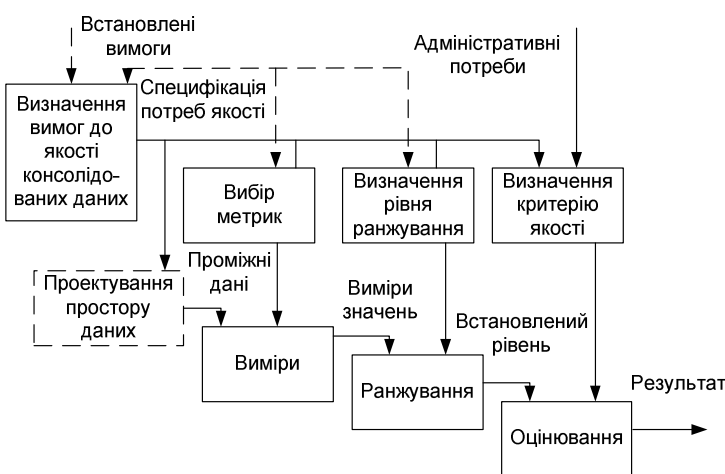


Рис. 6. Схема процесу оцінювання якості консолідованих даних

до складу простору даних, є незалежними. Тому, в першу чергу, досліджуватиметься релевантність знайдених даних та їх вклад у якість прийняття рішень на їх основі. Процес оцінювання якості консолідованих даних складається з 3-х стадій: встановлення вимог до якості консолідованих даних, підготовка до оцінювання та процедура оцінювання (рис. 6).

Поряд з фактичною оцінкою якості консолідованої інформації (10) необхідно провести оцінку якості еталонного зразка (11), що відображає найкраще прийняте рішення. Потім проводиться нормування фактичної оцінки за формулою (12).

$$K_i = \sum_i \alpha_i M_i, \quad K'_{const} = \sum_i q_i K, \quad (10)$$

$$K_i^e = \sum_i \alpha_i M_i^e, \quad K'_{const}{}^e = \sum_i q_i K^e, \quad (11)$$

де α_i – ранг важливості, m_i – відносне значення i -ї характеристики, k_i – значення i -о показника якості, що визначається експертно, причому $k_i \in [0; 1]$,

$$K_{cons} = K'_{cons} / K^e_{cons}, \quad (12)$$

Далі розробимо алгоритм визначення відповідності рішення, прийнятого на основі консолідованих даних, еталонному.

4.5. Розроблення алгоритму визначення відповідності рішення еталонному

Укрупнений алгоритм визначення відповідності рішення еталонному подано так.

1. Отримання параметрів вибірки еталонних та консолідованих даних.
 2. Визначення критеріїв оптимальності.
 3. Визначення найкращого значення за критерієм.
 4. Визначення найгіршого значення за критерієм.
 5. Пошук прямо пов'язаних даних (через відношення Dic).
 6. Групування вибраних даних.
 7. Вибір тих консолідованих даних, у яких агреговані кількісні характеристики рівні середньому значенню критеріїв 2) і 3).
 8. Визначення джерела даних, з якого отримано інформацію, що задовольняє п. 7.
- Ступінь співпадіння критерію з еталонним для заданих параметрів буде визначатись як

$$s = \sum_{i=1}^n a_i + a_t, \quad (13)$$

де n – кількість нечасових параметрів співставлення;

a_i – значення нечасового i -го параметра співставлення, яке набуває значення

$$a_i = \begin{cases} 0, & \text{параметр невідмічений} \\ 1, & \text{параметр відмічений} \end{cases};$$

a_t – значення часового параметра, яке визначається як

$$a_t = \begin{cases} 0, & \text{ігнорувати дати плану} \\ 1, & \text{ігнорувати дати документа} \\ 2, & \text{не враховувати інтервали плану} \\ 3, & \text{перетин інтервалів} \\ 4, & \text{повне співпадіння інтервалів} \end{cases}.$$

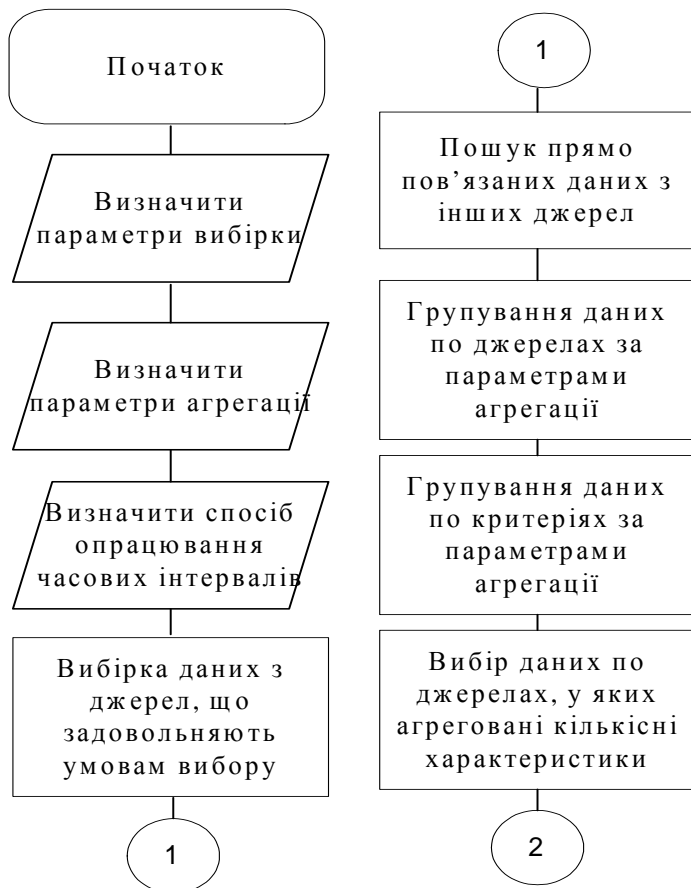


Рис. 7. Алгоритм визначення відповідності прийнятого рішення еталонному (початок)

Блок-схему алгоритму визначення ступеня співпадіння подано на рис. 7 – 8.



Рис. 8. Алгоритм визначення відповідності прийнятого рішення еталонному (закінчення)

Слід зазначити, що кількість параметрів співставлення n буде різною для кожного типу рішення, що приймається. Вибір того чи іншого параметра фізично означатиме, що при співставленні знайдених даних та еталона за обраним атрибутом буде виконуватись операція агрегації для визначення релевантності. Чим більше параметрів буде включено до агрегування, тим точнішим буде отриманий результат співставлення. Вибір усіх параметрів означає максимальний ступінь довіри до отриманих результатів співставлення.

Можуть бути отримані такі результати співставлення v :

- еталон не має аналога, $v = 0$;
- знайдені дані не відповідають жодному еталону, $v = 1$;
- часткове співпадіння (при агрегації даних еталона та знайдених даних отримано кількісні характеристики, які не рівні між собою), $v = 2$;
- повне співпадіння, $v = 3$.

4.6. Визначення функціональної придатності даних простору даних

Коректність або вірогідність даних – це ступінь відповідності даних про об'єкти в СД або ПД реальним об'єктам у момент часу, що визначається змінами самих об'єктів, некоректностями записів про їхній стан або некоректностями розрахунків їхніх характеристик. Вибір і встановлення вимог до коректності даних можна оцінювати за показниками накопичення, актуальності та достовірності даних. Сюди ж можна віднести й деякі часові характеристики даних та характеристики обсягу (рис. 9):

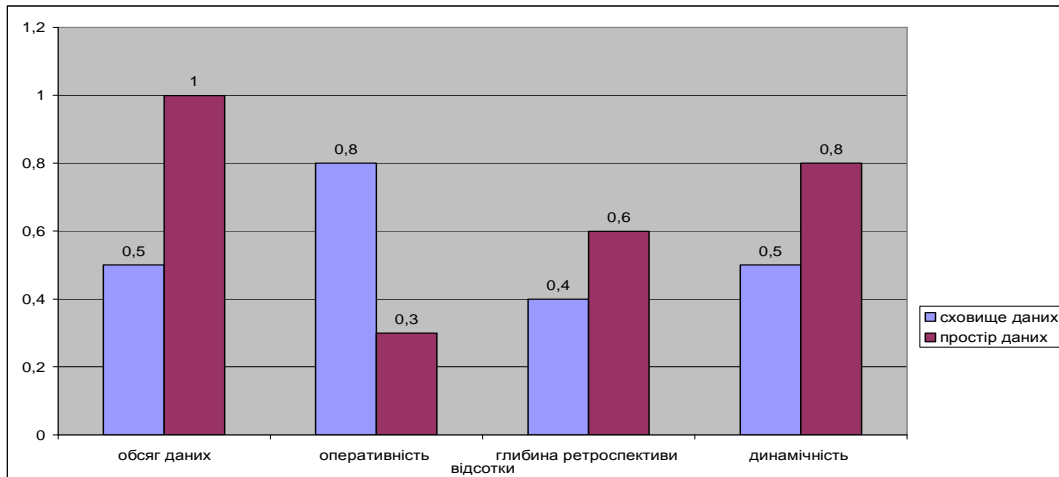


Рис. 9. Порівняння коректності сховища та простору даних

1. Обсяг даних – відносна кількість записів про об'єкти або документи, що доступні для зберігання й опрацювання, у порівнянні з кількістю реальних об'єктів у зовнішньому середовищі.
2. Оперативність – ступінь відповідності динаміки зміни даних станам реальних об'єктів.
3. Глибина ретроспективи – інтервал часу від найменшої дати випуску й/або запису документа дотепер.
4. Динамічність – відносна кількість описів об'єктів, що змінили свої характеристики, до загальної кількості записів за деякий інтервал часу.

4.7. Простір даних мережі супермаркетів

Для апробації методу було розв'язано задачу консолідації даних у режимі реального часу з робочих місць касирів супермаркетів, які розділені географічно.

Однією з проблем, що виникала з використанням традиційних способів завантаження (2 рази на день з припиненням роботи усіх кас), було надмірне навантаження на сервер. Окрім того, тактичні рішення – довозення продукції, якої зараз бракує, ускладнювалися у зв'язку з недостовірною інформацією.

Стратегічні рішення щодо керування мережею супермаркетів приймаються на основі консолідованих даних і зараз немає потреби у надоперативності даних. Проте керівництво вирішило перевірити, наскільки якість рішення буде залежити від швидкості потрапляння даних до центрального сховища.

Для визначення якості консолідованих даних проаналізуємо, наскільки співпадають прийняті рішення щодо стратегічного розвитку на основі відомих (еталонних) даних, консолідованих даних та даних, отриманих традиційним шляхом (через формування статистичних звітів).

Було визначено такі показники:

- 1) розмір інвестицій у персонал;

- 2) розмір інвестицій у розвиток транспорту;
- 3) розмір інвестицій у збільшення асортименту продукції;
- 4) розмір інвестицій у відкриття нових супермаркетів;
- 5) розмір інвестицій у програмне забезпечення;
- 6) розмір інвестицій у розвиток інфраструктури;
- 7) розмір інвестицій у рекламу;
- 8) розмір інвестицій в управлінську сферу;
- 9) розмір інвестицій в енергозберігаючі установки.

Результат порівняння реальних показників та показників по консолідованих даних подано на рис. 10.

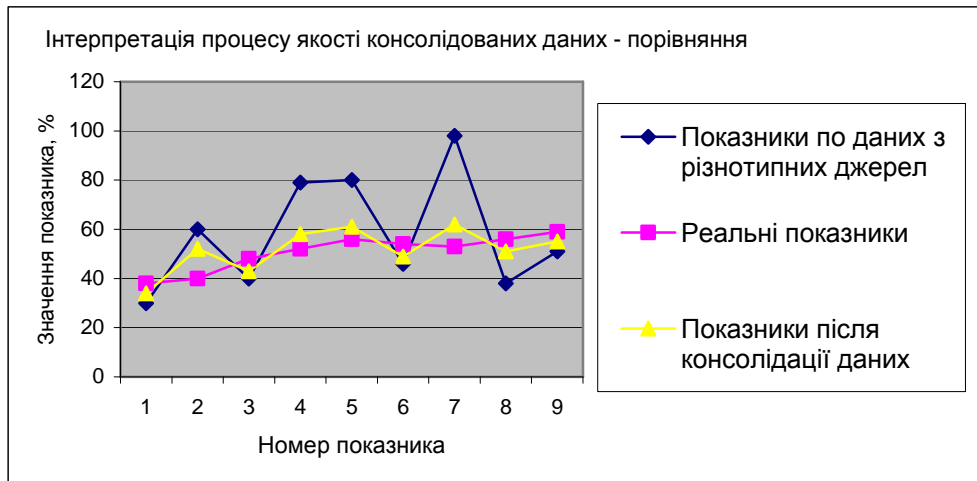


Рис. 10. Дослідження якості прийняття рішень по консолідованих даних

5. Висновки

У роботі розроблено метод визначення якості консолідованих даних, що дозволило визначити придатність цих даних для подальшого прийняття рішень. Науковою новизною роботи є формалізація характеристик якості програмних продуктів, описаних в ISO 9126. Розроблено метод визначення якості консолідованих даних, що дозволило визначити придатність цих даних для подальшого прийняття рішень. Цей метод є основою для методів і засобів побудови просторів даних як інформаційних продуктів. Практичною цінністю є розроблення алгоритму визначення якості консолідованих даних.

СПИСОК ЛІТЕРАТУРИ

1. Рогушина Ю.В. Формирование тезауруса предметной области как средства моделирования информационных потребностей пользователя при поиске в Интернете / Ю.В. Рогушина, А.Я. Гладун // Вестник компьютерных и информационных технологий. – 2007. – № 1. – С. 26 – 33.
2. ANSI/NISO Z39.88-2004. The OpenURL Framework for Context-Sensitive Services. – National Information Standards Organization [Електронний ресурс]. – 2005. – Режим доступу: http://www.niso.org/standards/resources/Z39_88_2004.pdf.
3. Инициатива SDMX: новые подходы к обмену статистическими данными [Электронный ресурс]. – Режим доступа: <http://citforum.univ.kiev.ua/internet/xml/sdmx>.
4. Калиниченко Л.А. СИНТЕЗ – язык определения, проектирования и программирования интероперабельных сред неоднородных информационных ресурсов / Калиниченко Л.А. – М.: ИПИ РАН, 1993. – 115 с.
5. Kalinichenko L.A. Compositional Specification Calculus for Information Systems Development // Proc. of the East-West Conference on Advances in Databases and Information Systems (ADBIS'99), (Maribor, Slovenia, September 1999). – Maribor, Slovenia: Springer Verlag, LNCS, 1999. – P. 317 – 331.

6. Ступников С.А. Формальные методы и модели в композиционных инфраструктурах распределенных информационных систем // С.А. Ступников // Системы и средства информатики. Специальный выпуск. – М.: ИПИ РАН, 2005. – С. 40 – 68.
7. Кулик Б.А. Представление логических систем в вероятностном пространстве на основе алгебры кортежей. 1. Основы алгебры кортежей / Б.А. Кулик // Автоматика и телемеханика. – 1997. – № 1. – С. 126 – 136.
8. Dong X. Indexing Dataspaces / X. Dong, A. Halevy // SIGMOD'07. – Beijing, China, 2007. – June 11–14. – P. 32 – 41.
9. Lenzerini M. Data Integration: A Theoretical Perspective [Электронный ресурс] / М. Lenzerini // PODS. – 2002. – P. 233 – 246. – Режим доступа: <http://www.dis.uniroma1.it/~lenzerin/homepage/talks/TutorialPODS02.pdf>.
10. Целишев А.С. Модификация метода критических работ с учетом динамически изменяемого состава ресурсов в распределенных вычислениях // А.С. Целишев, В.В. Топорков // Труды XV междунар. научно-техн. конф. «Информационные средства и технологии»: в 3 т. – М.: Издательский дом «Мэи», 2007. – Т. 1. – С. 340 – 345.

Стаття надійшла до редакції 04.03.2011