

КОНЦЕПТУАЛЬНІ ОСНОВИ СТВОРЕННЯ БАЗИ ДАНИХ НАУКОВОГО ЕКСПЕРИМЕНТУ ТА СПОСТЕРЕЖЕННЯ

*Інститут ядерних досліджень НАН України, Київ, Україна

Анотація. У статті наведено підхід до створення спеціалізованих баз даних, призначених для впорядкування та збереження значних обсягів інформації, що використовується в ході наукового спостереження та експерименту. Наведено функції і предметну область такої бази даних. Представлено модель об'єкта спостереження й зазначено задачі класифікації. Розроблено концептуальну модель бази даних експерименту. Представлено приклад можливої реалізації бази даних фізичного експерименту, яка зараз знаходиться у стадії розробки. Зроблено висновки.

Ключові слова: база даних, класифікація, множина, змінні, модель.

Аннотация. В статье приведен подход к созданию специализированных баз данных, предназначенных для упорядочения и сохранения значительных объемов информации, используемой в ходе научного наблюдения и эксперимента. Представлены функции и описана предметная область такой базы. Описано построение модели объекта наблюдения и указаны задачи классификации. Разработана концептуальная модель базы данных эксперимента. Приведен пример возможной реализации базы данных физического эксперимента, которая сейчас находится в стадии разработки. Сделаны выводы.

Ключевые слова: база данных, классификация, множество, переменные, модель.

Abstract. The paper presents an approach to the creation of specialized databases oriented to regulate and save significant amounts of information used in the course of scientific observation and experiment. The functions of such database were presented. The subject area database was described. Building a model of the object of observation was presented. The problems of classification were indicated as well. Conceptual database model of the experiment was designed. The example of a possible implementation of the database physical experiment was shown. This database is now under development. The conclusions were done.

Keywords: database, classification, set, variables, model.

1. Вступ

Закон України «Про наукову і науково-технічну діяльність» від 26.11.2015 № 848 націлений на створення умов для підвищення ефективності наукових досліджень і використання їх результатів для забезпечення розвитку всіх сфер суспільного життя. У ст. 1 вказаного Закону визначається термін «дослідницька інфраструктура» як сукупність засобів, ресурсів та пов'язаних з ними послуг, які використовуються науковим співтовариством для проведення досліджень на найвищому рівні, що охоплює найважливіші об'єкти наукового устаткування та обладнання або набори приладів, ресурси, що базуються на знаннях, до яких, зокрема, віднесені різноманітні бази та банки даних наукової інформації.

Проте, дійсний стан справ з формуванням інформаційних ресурсів у більшості наукових установ, не пов'язаних з розробками у сфері інформаційних технологій, дозволяє зробити такі висновки:

– дослідники не створюють і не працюють з базами і банками даних, інформація за значною частиною наукових експериментів та досліджень знаходиться, у найкращому випадку, у вигляді текстових документів Word або таблиць Excel, у найгіршому – в паперовому варіанті;

– при використанні програм обробки даних остаточні результати зберігаються у форматах цих програм, які частково трансформуються у загальноприйняті формати під час

роботи над звітами чи статтями, проте більшість інформації залишається у первинному варіанті, місце зберігання якої часто відоме лише фахівцю, який обробляв ці дані;

– для того, щоб знайти інформацію про минулі дослідження чи результати експериментів, які були проведені, слід піднімати паперові звіти минулих років, за необхідності їх відцифрувати самостійно, проте, зважаючи на стан архівних документів, багато інформації при цьому втрачається.

Зазначене іноді стає на заваді міжнародному співробітництву вчених, адже широке використання баз даних у дослідницьких організаціях за кордоном потребує відповідної структури організації даних і інформації від вітчизняних науковців.

Актуальність теми даної статті полягає в необхідності створення баз даних наукових експериментів та спостережень, які дозволять за допомогою сучасних засобів обчислювальної техніки працювати віддалено, паралельно, мінімізувати витрати часу та вірогідність помилки введення за допомогою встановлення відповідних ідентифікаторів і кодування, використання електронних паспортів речовин, лабораторій, окремих одиниць обладнання, територій, де проводяться експерименти, тощо.

Метою роботи є визначення концептуальних основ створення баз даних наукових експериментів та спостережень для підвищення контролю за дотриманням алгоритму дій, що моделюються в ході експерименту, а також рівня оперативності обробки даних та організації роботи з великими обсягами інформації.

Завдання роботи:

– дослідити функції та предметну область баз даних, призначених для наукового експерименту і/або спостереження;

– навести можливе представлення моделі об'єкта, що досліджується, і задач класифікації результатів;

– представити концептуальну модель бази даних, призначену для збереження даних і інформації експерименту та продемонструвати її на прикладі створюваної бази даних для робіт, що проводяться в лабораторії фізико-технічних проблем джерел ядерних випромінювань відділу ядерної фізики (ЛФТПДЯВ ВЯФ) Інституту ядерних досліджень НАН України.

Дослідження тематики створення та використання банків і баз даних триває з часів впровадження перших електронних обчислювальних машин і були широко описані в роботах В.М. Глушкова [1]. Сучасні відомі дослідники К. Дж. Дейт [2], Т. Конноллі [3], Г. Гарсія-Моліна [4] та ін. представляють підходи до проектування, створення та використання баз даних у різних сферах практичної діяльності. В мережі Інтернет наведено декілька прикладів реалізації за допомогою Web-інтерфейсу баз даних спостережень за птахами, погодою [5–7]. Бази даних наукових експериментів, як правило, вузькоспеціалізовані, створюються під конкретну задачу, представлені таблицями отриманих даних та візуалізацією результатів програм обробки з можливістю віддаленої роботи та управління інтерфейсом користувача [8].

2. Функції та предметна область бази даних наукового спостереження та експерименту

Ще у минулому столітті безперервне зростання обсягів інформації привело до створення програмних комплексів, що були названі СУБД – системи управління базами даних. СУБД не лише визначали процедури введення та збереження даних, але й дозволили описати структуру інформації, що і визначило сутність та перспективи розвитку баз даних (БД).

Визначення наукового експерименту представляє цю категорію як метод наукового дослідження, що полягає у відтворенні процесів або явищ у лабораторних умовах з накопиченням відповідної інформації [9]. Спостереження є безпосереднім способом отримання дослідних даних. Враховуючи зазначене, база даних наукового експерименту та спостере-

жень буде включати велике число компонентів, об'єднаних між собою складною системою функціонального зв'язку та інформаційного обміну. Ці компоненти повинні бути строго регламентовані та узгоджені у часі, а також підпорядковані загальному алгоритму проведення експерименту або спостереження з можливістю адаптації під цей алгоритм.

Є два шляхи реалізації такого підходу. Перший підхід – класичний – полягає у створенні інтелектуальної інформаційної технології з отримання, обробки і систематизації інформації в режимі реального часу, заснованій на спільному використанні методів інженерії знань і традиційних методів моделювання, оптимізації і роботи з базами даних. Доцільність такої інтеграції обумовлюється необхідністю використання при вирішенні практичних завдань обробки результатів випробувань як чисельних, так і логічних методів роботи з даними і знаннями.

Другий підхід – створення для реалізації задач наукового дослідження розподіленої інформаційної системи, орієнтованої на збір, зберігання, пошук і обробку текстової або фактографічної інформації, здатної функціонувати в умовах дії випадкових факторів за активної взаємодії із зовнішнім середовищем, за наявності негативних впливів різної природи. В основу такої системи закладається ідея організаційного та фізичного розподілу користувачів, що одночасно взаємодіють з базою даних, а також взаємоузгоджених, але розподілених логічно і фізично даних, що утворюють єдине ціле. Один із варіантів втілення такого підходу – побудова баз даних на основі онтологій. В основі онтологічної системи знань лежить об'єктно-орієнтований підхід, за яким прикладна область описується як сукупність об'єктів, що взаємодіють через трансляцію повідомлень.

Щоб обрати шлях реалізації бази даних наукового експерименту, слід визначити основні функції такої бази. У загальному випадку база даних для фізичного наукового експерименту чи спостереження повинна підтримувати такі основні функції:

- отримання інформації про хід випробувань від системи збору і передачі даних;
- ведення бази даних у реальному часі, відображаючи динаміку поведінки спостережуваного об'єкта;
- аналіз результатів спостереження на предмет відхилення параметрів, що контролюються, від номінальних значень (сюди ж може входити виявлення і прогнозування несприятливих тенденцій) і оперативне вирішення задач класифікації, розпізнавання, технічної діагностики та ін.;
- формування звітів з використанням, за необхідністю, методів моделювання, чисельної оптимізації і логічних висновків.

Концептуальна модель бази даних наукового експерименту чи спостереження – це віддзеркалення предметної області, для якої розробляється база даних. Наприклад, для фізичного експерименту база даних виступає системою контролю за дотриманням алгоритму дій, що моделюються в ході експерименту. Процедура контролю полягає в перевірці відповідності якості об'єкта певним вимогам, які задаються зазвичай у вигляді обмежень на показники властивостей. Якщо вони доступні для виміру та спостереження, тоді вони виступають як ознаки для визначення стану процесу експерименту. Такими обмеженнями в експерименті можуть виступати параметри або функції. Зазвичай кожній ознаці задаються межі її величин, у відповідності з якими співвідноситься, наприклад, вид технічного стану досліджуваного об'єкта. Проте в більшості випадків є істиною їх однозначна відповідність.

Процес контролю у фізичному експерименті в загальному випадку є подачею на об'єкт певних дій (вхідних сигналів) і багатократним виміром і аналізом реакцій (у відповідь вихідних сигналів). Дії на об'єкт дослідження поступають від засобів, що контролюються, або у вигляді зовнішніх дій (сигналів), визначених робочим алгоритмом функціонування. Тому в базу даних повинна поступати інформація про тестовий і функціональний контроль, а порівняння еталону з об'єктом, що досліджується, вимагає застосувати класифікацію множини об'єктів, що досліджуються у процесі експерименту.

У більшості випадків класифікація ґрунтується на деякій сукупності ознак (довіднику, словнику, тезаурусі), що описують кожен об'єкт дослідження. Вважається, що ознаки малочутливі до зміни об'єктів, які належать до одного класу. Обсяг довідника визначається достатнім для правильної класифікації. Визначають множину класів $\{E_i\}_{i=1}^m$. Властивості кожного класу E_i описуються множиною змінних $\{e_{ij}\}_{j=1}^n$.

Тоді об'єкт, який повинен бути розпізнаний та розміщений у базі даних, може бути охарактеризовано набором змінних $y_j (j = \overline{1, n})$, які отримуються з визначення ознак за довідником, що використовується. Класифікація полягає у тому, щоб співвіднести кожен новий об'єкт до одного з класів E_i . Рішення щодо належності об'єкта до класу приймається на основі розрахунку показника схожості $d(e, E_i)$ між вектором y_i та $\{e_{ij}\}_{j=1}^n$.

Слід зазначити, що кожен клас може бути описаний кількісними параметрами, багатовимірними функціями, логічними виразами, якісними показниками, густиною розподілу ймовірності ознак. Тоді показник $d(e, E_i)$ буде розраховуватися за різними підходами. Наприклад, як середньоквадратична відстань між y_i та сукупністю $\{e_{ij}\}$, через функцію ризику, ігрові методи та ін. Але при визначенні $\{e_{ij}\}$ логічними і якісними змінними єдиним можливим методом виступатиме метод логічного висновку.

3. Модель об'єкта і задачі класифікації результатів

Модель будь-якого об'єкта може бути представленою впорядкованою множиною:

$$\Delta = \langle T, X, Y, Z, F, L \rangle, \quad (1)$$

де T – множина моментів часу, в період яких спостерігається об'єкт;

X, Y – множина вхідних і вихідних параметрів відповідно;

Z – множина станів об'єкта;

F – оператор переходів, який відбиває механізм зміни станів об'єкта під впливом зовнішніх і внутрішніх факторів;

L – оператор виходів, який описує формування параметрів виходу як реакції об'єкта на вплив дії зовнішніх і внутрішніх факторів.

Оператори F та L реалізують відображення:

$$F : T \times X \times Z \rightarrow Z, \quad (2)$$

$$L : T \times X \times Z \rightarrow Y. \quad (3)$$

Будь-який стан об'єкта $z \in Z$ може бути представлений у будь-який момент часу $t \in T$ набором змінних $z_r (r = \overline{1, k})$, які мають різні значення під впливом зовнішніх і внутрішніх факторів. За підсумками результатів співставлення всіх змінних стану об'єкта зі значеннями, що задані апріорно і які характеризують рівень роботи об'єкта або змін, що відбуваються в результаті проведення дослідів, стан об'єкта може бути віднесено до якоїсь класифікації. Але таке співставлення не завжди можна здійснити, оскільки змінні стану $z_r (r = \overline{1, k})$ є абстрактними, фізична природа яких не завжди відома, тож і їх вимір не завжди можливий. На відміну від них вихідні змінні $y_j (j = \overline{1, n})$ можна спостерігати і ви-

мірювати, оскільки вони є конкретними фізичними величинами (сила току, напруження, кутові та лінійні зміщення та ін.). У цьому випадку вихідні показники більш зручні для завдань класифікації та фіксації у базі даних як ознак, які контролюються. Тобто, визначення виду технічного стану об'єкта можна здійснити на практиці не у просторі змінних стану $z_r (r = 1, \bar{k})$, а у просторі вихідних змінних $y_j (j = 1, \bar{n})$ або інших змінних, що є конкретними фізичними величинами.

З математичної точки зору, визначення будь-якого стану об'єкта можливе лише в тому випадку, коли за підсумками вимірів вихідних параметрів $y_j (j = 1, \bar{n})$ при відомих значеннях параметрів входу $x_s (s = 1, \bar{l})$ може бути отримана оцінка будь-якої зі змінних стану $z_r (r = 1, \bar{k})$. Тобто, вирішена задача спостереження.

Ця задача полягає у такому: на основі відомого вихідного процесора $y(t) \in Y$ визначити невідомий стан об'єкта $z(t) \in Z$, де $y(t)$ та $z(t)$ – вектор-функції. Формально задача полягає у вирішенні відносно $z(t)$ рівняння

$$L[t, x(t), \hat{z}(t), \tau] = \hat{y}(t), \quad (4)$$

де $\hat{y}(t)$ – опис деякої частини процесу на виході, що можливо зареєструвати за допомогою вимірювальних пристроїв.

Приймаємо, що об'єкт можна спостерігати у стані $z(t) \in Z$ на множині моментів часу $\hat{T} = \{\tau\}, \hat{T} \subset T$ при впливі на виході $x(t) \in X$, якщо рівняння (4) має єдине рішення:

$$\hat{z}(t) = z(t) \in Z.$$

Якщо твердження $z(t) \in Z \subset R^k$, то об'єкт знаходиться повністю під спостереженням і всі вхідні і вихідні параметри об'єкта на визначених часових проміжках вносяться у базу даних. Необхідна і достатня умова повного спостереження об'єкта означає, що будь-якій зміні вектора виходу $y(t) \in Y$ при фіксованому векторі $y(t) \in X$ відповідає певна зміна вектора стану об'єкта:

$$y(t_1) \neq y(t_2) \Rightarrow z(t_1) \neq z(t_2); \quad \forall t_1, t_2 \in T.$$

Завдяки цьому вхідні змінні $y_j (j = 1, \bar{n})$ можна використовувати як ознаки поточного стану об'єкта, який спостерігається. Це є результат першої задачі спостереження, який фіксується в базі даних наукового експерименту.

Повне спостереження за об'єктом може бути досягнуте за допомогою вибору точок контролю, де відбуватиметься збір сигналів виходу з передачею цієї інформації до бази даних. Тобто, вибір таких точок є найбільш важливим моментом спостереження або експерименту. У цьому випадку задачею класифікації є віднесення стану об'єкта, що спостерігається у будь-якій з точок, до заданого виду об'єкта. Вирішення цієї задачі є другим результатом процесу контролю здійснення експерименту, який заноситься до бази даних. Вирішенням цієї задачі є відшукання відображення

$$\eta: Y \rightarrow E, \quad (5)$$

де E – множина заданих видів стану об'єкта.

Задача класифікації полягає у розбитті множини Y на ряд класів, що не перетинаються, і у визначенні належності кожного стану об'єкта, що спостерігається, до одного з

класів. Множину можна розглядати як модель фактор-множини класів Y/Q , які не перетинаються. Вирішення цих питань зводиться до побудови моделі об'єкта, що контролюється, враховуючи реальні умови його функціонування і вимоги, які висуваються до отриманих результатів.

Третьою задачею класифікації є реалізація відображення

$$\Psi: E \rightarrow S, \quad (6)$$

яке є формальним у відношенні до визначеного виду стану об'єкта, що досліджується, та наводить відповідність щодо рішення про стан об'єкта з урахуванням імовірнісних характеристик можливих помилок, похибок, перешкод.

4. Концептуальна модель створення бази даних експерименту

Усі перелічені задачі класифікації об'єкта дослідження для створення бази даних наукового спостереження або експерименту можуть бути представлені узагальненою моделлю (рис. 1).

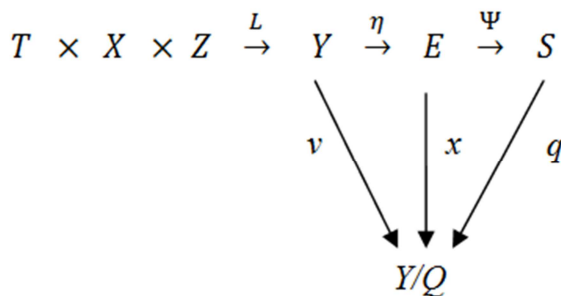


Рис. 1. Узагальнена модель бази даних наукового експерименту

Реалізація процесу наукового спостереження при реалізації експерименту згідно з наведеною моделлю полягає у виконанні двох етапів. На першому етапі відбувається велика кількість різноманітних спостережень, пов'язаних з імітацією різноманітних станів об'єкта, у тому числі – відмови, спостереження в лабораторних чи природних умовах. До таких спостережень належать і дослідження реально існуючих об'єктів. При кожному відтворенні чергового стану об'єкта послідовно реалізуються відображення L та v . Цей етап є трудомістким,

пов'язаний зі значними витратами часу та ресурсів, але він лежить в основі другого етапу – визначення фактор-множини станів Y/Q об'єкта дослідження та множини видів його технічного стану E , який знаходиться у взаємно однозначному відношенні до Y/Q , з урахуванням імовірнісних характеристик q можливих помилок, похибок, перешкод.

З погляду на весь цикл можливих ситуацій під час спостереження за об'єктом, може виникати ситуація, коли відбуваються події або процеси, виникнення яких не передбачено алгоритмом експерименту. Ці події або процеси слід внести до бази даних у режимі реального часу для роботи інших фахівців з метою відповіді на питання: чому процес пішов іншим шляхом або відбувається нерегламентована подія. У цьому разі показник управління в реальному часі T можна визначити як $T = 0$ – ознака своєчасності інформації, що внесена до бази даних; $T = 1$ – ознака несвоечасності внесення інформації. Таку ознаку отримує кожний запис до бази даних системи обробки інформації, в якому виявлено несвоечасність. Для здійснення такої перевірки кожний запис має містити в собі визначений інтервал часу, коли інформація в записі є актуальною. У загальному вигляді його можна визначити як додаткову залежність інформації, що отримана в ході експерименту, від відрізка часу, на якому вона є актуальною:

$$T = 0, \text{ якщо } \begin{cases} x_s(a, b, c, d, t_0 - t_k) \in X, \\ y_j(a, c, d, e, t_0 - t_k) \in Y; \text{ якщо } t_0 < t < t_k, \\ z_r(a, c, d, e, t_0 - t_k) \in Z, \end{cases} \quad (7)$$

де x_s, y_j, z_r – визначені вище, a – ім'я події чи процесу, що досліджується, b – кількість точок спостереження, c – належність до класу, d – ознаки цього класу, e – ознаки операції, що проводиться в рамках експерименту.

Додатково до наведених (7) у записах бази даних ураховуються ознаки актуальності запису у часі на інтервалі $t_0 < t < t_k$, який визначається алгоритмом експерименту. Тобто, запис у базі даних про якусь подію є актуальним для дослідження, вивчення, формування звіту, бо співпадає з часовим інтервалом адекватності інформації.

Вихід записів x_s, y_j, z_r за межі визначеного відрізка часу робить інформацію архівною. У той же час вихід за межі визначеного відрізка часу інформації про вихідні параметри z_r є показником реалізації алгоритму експерименту, оскільки визначає факт отримання даних на виході незалежно від того, успішним чи неуспішним був експеримент.

5. Створення бази даних фізичного експерименту

Враховуючи викладене, база даних наукового експерименту чи спостереження повинна, з одного боку, утримувати всі властивості класичної бази даних, з іншого, мати особливості розподіленої інформаційної системи з активною взаємодією із зовнішнім середовищем та стійкою до впливу випадкових факторів. Така база даних повинна бути реалізована у двох площинах: внутрішній (сервер) та зовнішній (Інтернет-портал). Це дозволить накопичувати й зберігати навіть первинну, необроблену, інформацію про всі існуючі експерименти та спостереження на сервері, а вибірково інформацію, необхідну для роботи більшості дослідників, викладати з обмеженим чи частково обмеженим доступом на Інтернет-портал. Крім того, через Інтернет-портал можна організувати доступ он-лайн до заповнення та коригування таблиць проведення експерименту (спостереження), а також до процесу обробки інформації, зібраної учасниками експерименту (статистична обробка, виявлення ознак класифікації та ін.).

Розглянемо приклад бази даних, що знаходиться у процесі розробки для виконання експериментів за тематикою досліджень міграції тритію в оточуючому середовищі. Для реалізації бази даних обрана СУБД Oracle.

Експеримент передбачає не лише збір та аналіз даних про переміщення радіонуклідів водню – тритію – в оточуючому середовищі і на цій основі створення моделі на ділянці натурних експериментів. У процесі виконання роботи передбачається дослідження та відтворення можливих бар'єрних методик, експерименти розповсюдження тритію за харчовими ланцюгами та ряд дослідів з проникнення тритію через різноманітні матеріали. Все це в підсумку повинно дозволити розробити заходи з мінімізації негативного впливу тритію на довколишнє середовище та організм людини шляхом забезпечення води як речовини, що легко заміщує атом водню на атом тритію та складає 80% вмісту живих організмів. З цією метою ряд робіт з дослідження будуть вестися паралельно, дані необхідно отримувати в режимі реального часу з лабораторії та ділянки натурального експерименту. Одночасно з базою даних працюватимуть від 5 до 7 науковців.

Орієнтуючись на задачі експерименту, можна зазначити, що представлене моделью (4) – (7) є вимогами щодо виконання конкретних обов'язків щодо вирішення задач класифікації учасниками експерименту. Узагальнені задачі експерименту (рис. 1) утримують опис прецедентів, вказівку на використання конкретних даних та послідовність виконання

процесів. Тож усі системні прецеденти можна представити на мові графічного опису для об'єктів моделювання у сфері розробки програмного забезпечення (UML – Unified Modeling Language) у вигляді діаграми прецедентів (рис. 2).

На рис. 2 суцільними стрілками представлено потік дій управління, переривчастими – виконання дій над об'єктами. Товста суцільна стрілка вказує послідовність дій з наповнення бази даних та формування підсумкового звіту. «Перевірка відповідності_1» є перевіркою отриманої інформації на відповідність поставленим задачам експерименту. «Перевірка відповідності_2» є перевіркою обробленої інформації на відповідність меті експерименту. Саме ця інформація, за задумом, і повинна транслюватися через Інтернет-портал для інших користувачів, що напряду не задіяні у формуванні бази даних, але використовують дані експерименту для своєї роботи.

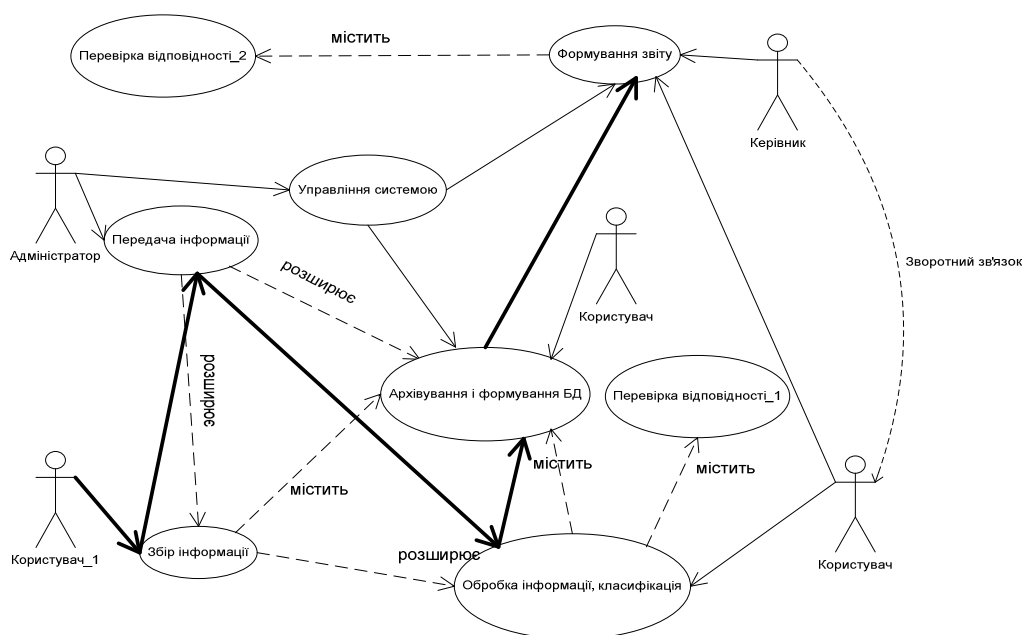


Рис. 2. UML-діаграма системних прецедентів збору, обробки та передачі інформації для розміщення у базі даних

В UML-діаграмі зазначені два спеціальних типи зв'язку між прецедентами:

- «містить» – коли один прецедент у процесі свого виконання обов'язково здійснює деяку послідовність дій, які входять до іншого прецеденту;
- «розширює» – коли прецеденти подібні за своїми діями, але один із них несе дещо збільшене функціональне навантаження.

На основі системних прецедентів можна побудувати UML-діаграму класів. Найменуваннями класів будуть виступати окремі види робіт, які несуть у собі чітко визначене функціональне навантаження по збору інформації, її обробці та формуванню бази даних. Це дозволить виконати всю систему збору, обробки та архівування інформації у вигляді модулів, що є основою гнучкої інтеграції та легкої адаптації до умов окремих експериментів. Особливістю діаграми класів для бази даних зазначеного експерименту виступатимуть три метакласи, які повинні утримувати вхідну інформацію, на основі якої реалізується алгоритм експерименту:

- метаклас «Паспорт речовини» забезпечує інформацією про технічні дані хімічної речовини, металу, сплаву, які використовуються в експерименті. Так, наприклад, для тритію електронний паспорт речовини буде представлений позиціями: апарат (контейнер зберігання), кількість у літрах та за випромінюванням, повнота насичення, ізотопний склад, упа-

ковка, дата випуску, дата збереження. Інші речовини можуть бути представлені позиціями, що найбільш точно описують фізико-хімічні властивості речовини;

– метаклас «Паспорт лабораторії» уміщує в собі регламентовані дані та внутрішні нормативні документи на лабораторію та на ділянку натурального експерименту. В паспорті представлена повна характеристика обладнання, що використовується в лабораторії та на ділянці натурних експериментів. Окрім того, у паспорті ділянки натурних експериментів описуються всі природні об'єкти, які входять до зони досліджень. Тут же використовується QR-маркування для автоматичного переходу на сторінку спостережень, розташованої на Інтернет-порталі, за кожним природним об'єктом з метою внесення оперативних даних: «Дерево_1» → «QR-код» → «Таблиця спостережень». Це дозволяє безпосередньо в режимі он-лайн вносити та коригувати дані, а також уникнути зайвої паперової роботи та похибки при перенесенні даних з паперового на електронний носій;

– метаклас «Забезпечення вхідною інформацією» дозволяє отримати дані у різних форматах безпосередньо з місця події, структурувати за певними ознаками та передавати для подальшої обробки або безпосередньо до архіву.

Для ведення і використання бази даних передбачаються обмеження доступу користувачів і постійний контроль роботи з боку системного адміністратора для забезпечення збереження даних і конфіденційності інформації. Для оперативної роботи деяким віддаленим користувачам створюється доступ на вхід до певної сторінки через QR-код ідентифікації користувача. Окрім того, QR-кодування використовується для доступу до літературних джерел за проектом.

Як зазначалося, класифікація ґрунтується на деякій сукупності ознак, представлених у довіднику з описом кожного об'єкта дослідження. Для бази даних використані системні довідники фізичних показників, дати, одиниць виміру, типів динаміки та джерел даних. Наприклад, довідник «Джерела даних» призначений для зберігання інформації про джерела даних показників, що використовуються в системі. Довідник застосовується в системі для забезпечення користувачів і розроблювачів системи інформацією про джерела даних за різними показниками і складається з полів: код, ідентифікатор елемента, найменування, опис, відповідність, розміщення (на Інтернет-порталі). Специфікація довідника «Джерела даних» наведена в табл. 1.

Таблиця 1. Специфікація довідника «Джерела даних»

Найменування	Тип даних	Not NULL	Primary	Foreign Key	Опис
Code	varchar(255)				Код
ElemID	int	X	X		Ідентифікатор елемента
Name	nvarchar(1024)	X			Найменування
Description	nvarchar(4000)				Опис
ChangeID	int				Ідентифікатор відповідності
SITE	varchar(256)				Сайт

Велика кількість різноманітних таблиць ускладнює задачу створення вибірки даних для підготовки звіту або розрахунку якоїсь моделі в рамках проведення дослідження. Для вирішення цієї задачі на СУБД Oracle застосовано підхід до організації зв'язку таблиць шляхом формування дочірніх таблиць для отримання вибірки даних. Наприклад, для створення моделі забруднення ділянки натурних експериментів техногенним тритієм одним із можливих і оптимальних варіантів побудови такого зв'язку буде виведення підсумків SQL-запису за допомогою PL/SQL. PL/SQL-запит може бути представлений таким кодом:

```
1) SELECT tt1.c1 AS tt1_c1, tt1.c2 AS tt1_c2, tt2.c1 AS tt1_c1, tt2.c2 AS tt2_c2, tt1_c1 AS tt2_tt1_c2;
```

- 2) FROM tt1;
- 3) left join tt2 ON tt1.c1 = tt2.tt1_c1.

Зв'язок бази даних СУБД Oracle та інших програмних виробів, які використовуються під час проведення наукового спостереження або експерименту й видають результати вимірів у вигляді таблиць, може бути проведений за допомогою стандартного засобу database link.

6. Висновки

У роботі розглянуто задачу створення бази даних наукового експерименту та спостереження, яка утримує всі властивості класичної бази даних і має особливості розподіленої інформаційної системи. Реалізація такої бази даних пропонується на внутрішньому сервері організації та Інтернет-порталі, через який відбуватиметься доступ користувачів до заповнення та коригування таблиць проведення експерименту або спостереження.

За підсумками роботи можна зробити такі висновки:

– для експерименту база даних виступає системою контролю за дотриманням алгоритму дій, що моделюються в ході експерименту, тож головною вимогою при створенні такої бази даних буде виступати адекватне представлення даних щодо відповідності об'єкта, що досліджується, певним вимогам показників властивостей;

– у базі даних експерименту повинні бути представлені три задачі класифікації стану об'єкта, що спостерігається, до заданого виду об'єкта. Це – вибір точок контролю, їх опис (відображення) та реалізація відображення, що є формальним по відношенню до визначеного виду стану об'єкта, що досліджується, й наводить відповідність щодо рішення про стан об'єкта з урахуванням імовірнісних характеристик можливих помилок, похибок, перешкод;

– концептуальна модель бази даних наукового експерименту чи спостереження охоплює збір, обробку (систематизацію) та архівування великих потоків інформації при виконанні двох етапів експерименту. Перший етап – спостереження імітації різноманітних станів об'єкта, що моделюється, та співвіднесення з реально існуючим об'єктом. Другий етап – визначення фактор-множини станів об'єкта дослідження та множини видів його технічного стану з урахуванням імовірнісних характеристик можливих помилок, похибок, перешкод.

Як приклад наведено підхід до створення бази даних, яка зараз розробляється в рамках фізичного експерименту за тематикою досліджень міграції тритію в оточуючому середовищі лабораторією фізико-технічних проблем джерел ядерних випромінювань Інституту ядерних досліджень НАН України.

Наведене в роботі може бути використане для розробки баз даних у різних сферах наукових досліджень.

СПИСОК ЛІТЕРАТУРИ

1. Глушков В.М. Основы безбумажной информатики / Глушков В.М. – М.: Наука, Гл. редакция физ.-мат. литературы, 1987. – 552 с.
2. Дейт К.Дж. Введение в системы баз данных / Дейт К.Дж. – [8-е изд.]. – М.: Вильямс, 2005. – 1328 с.
3. Коннолли Т. Базы данных. Проектирование, реализация и сопровождение. Теория и практика / Т. Коннолли, К. Бегг. – [3-е изд.]. – М.: Вильямс, 2003. – 1436 с.
4. Гарсиа-Молина Г. Системы баз данных. Полный курс / Гарсиа-Молина Г., Ульман Дж., Уидом Дж. – Вильямс, 2003. – 1088 с.
5. База данных «Онлайн дневники наблюдений» [Электронный ресурс]. – Режим доступа: <http://ru-birds.ru/index.php/en/sample-sites-2>.

6. Проект Турухтан. База данных учета птиц [Электронный ресурс]. – Режим доступа: <http://philomachus.ru/project>.
7. Веселов В.М. Язык описания гидрометеорологических данных для IBM PC совместимых ПЭВМ / В.М. Веселов // Труды ВНИИГМИ-МЦД. – 1995. – Вып. 160. – С. 41 – 54.
8. Редактор данных эксперимента [Электронный ресурс]. – Режим доступа: <http://www.ioffe.ru/LLT-SH/safon/ExpDataViewer.html>.
9. Ахутин А.В. Эксперимент и природа / Ахутин А.В. – СПб.: Наука, 2012. – 660 с.

Стаття надійшла до редакції 18.01.2016