контексти, в яких вони зустрічаються) зберігається у базі даних системи. Документи, завантажені та додані в базу документів системи, зберігаються в окремому каталозі.

У режимі консультації система, використовуючи дані, сформовані в процесі дослідження, надає консультації для користувачів системи (лінгвістів-перекладачів) щодо перекладу лексичних одиниць, їхнього тлумачення, етимології, а також подає цитати з реальних документів євроінтеграційного дискурсу, в яких використовується ця лексична одиниця.

Серед бажаних удосконалень цієї системи – підвищення рівня автоматизації відбору документів євроінтеграційного дискурсу. В реалізованому прототипі відбирає документи лінгвіст-дослідник на основі особистого досвіду. Удосконалення системи полягає у автоматичному відслідковуванні документів, які публікуються в Інтернеті, та визначення документів, які належать до євроінтеграційного дискурсу. Можлива і автоматизація пошуку самих неологізмів. Для реалізації описаних вдосконалень потрібно розв'язати ряд технологічних, алгоритмічних та емпіричних задач.

Програму було протестовано лінгвістом-експертом у галузі дослідження неологізмів євроінтеграційного дискурсу. Збоїв у роботі системи не виявлено.

1. Арутюнова Н.Д. Дискурс : лингвистический энциклопедический словарь. – М.: 1990. – С. 136–137. 2. Караулов Ю.Н., Петров В.В. От грамматики текста к когнитивной теории дискурса // Дейк Т. А. ван. Язык. Познание. Коммуникация. – М., 1989. – С. 8. 3. Dijk T.A. van. Discourse Studies. – London: Sage, 1997. – 245 p. 4. Карасик В.И. Языковой круг: личность, концепты, дискурс. – М.: Гнозис, 2004. – 390 с. 5. Зацний Ю.А. Сучасний англомовний світ і збагачення словникового складу. – Львів: ПАІС, 2006. – 228 с. 6. Литвин В.В., Даревич Р.Р., Досин Д.Г. Використання агент них систем, керованих онтологією, для пошуку інформації в мережі Інтернет // Вісн. Нац. ун-ту "Львівська політехніка". – 2007. – № 589. – С. 164–174. 7. Даревич Р.Р. Метод оцінювання подібності текстових документів, доповнених контекстом з онтології / Р.Р. Даревич, Д.Г. Досин, В.В. Литвин, Н.В. Никитюк // Відбір і обробка інформації. – 2007. – № 27(103). – С. 109–115.

УДК 004

Н.Б. Шаховська, О.А. Лозицький
Національний університет "Львівська політехніка",
кафедра інформаційних систем та мереж

ПРОСТОРИ ДАНИХ У ВИСТАВКОВІЙ ДІЯЛЬНОСТІ

Проаналізовано проблеми, що виникають під час роботи з розрізненими джерелами з використанням сховищ даних та баз даних. Уведено модель простору даних як засобу інтеграції та опрацювання даних з розрізнених джерел.

Problems which arise up during work with separate sources with the use of depositories information and databases are analysed. Described model of space of information as mean of integration and working of information from separate sources.

Вступ

За останні 10–15 років виставкова індустрія стала окремою галуззю. Адже до початку 90-х в Україні існувало лише два суб'єкти: Торгово-промислова палата та Виставка досягнень народного господарства УРСР, які організовували заходи в Україні і на теренах колишнього Радянського Союзу.

Відтак, на початку незалежності дуже важливу роль продовжували відігравати Торгово-промислові палати, що на початку були практично єдиними виставковими центрами. Згодом з'являються приватні професійні організатори виставок, а також різноманітні види виставкового сервісу.

За десять років незалежності виставкова діяльність в Україні набула сучасних розвинених форм, стала важливим чинником економічного розвитку. Були започатковані виставки практично з усіх економічно та соціально важливих виставкових тематик, сформувалося коло професійних виставкових компаній. Сьогодні у кожній галузі національної економіки сформувалися одна–дві провідні виставки, визначилися провідні організатори. Можна констатувати, що етап формування виставково-ярмаркової галузі загалом завершився.

У середині 90-х назріла потреба керувати процесами в галузі, забезпечувати галузеве координування задля сталого розвитку. У 1997 році була заснована Виставкова федерація України, яка на добровільних засадах об'єднує практично усіх серйозних учасників цього ринку. ВФУ – це професійне галузеве об'єднання, яке має на меті сприяти розвитку виставкової діяльності в Україні, а також – захищати законні права та інтереси її членів.

Виставкова федерація України є ініціатором багатьох важливих для галузі рішень. Галузеві стандарти, боротьба із недобросовісною конкуренцією, проблема прозорості і достовірності виставкової статистики, розвиток професійної освіти і трансфер технологій – це питання, які вирішує федерація.

Фірми та організації, які займаються виставковою діяльністю найефективніше, використовують переваги Інтернету для реклами та власне організації виставкових заходів. Оскільки зараз Інтернет є доступним практично для будь-якої людини, то можна констатувати той факт, що Інтернет став окремою, дуже важливою ланкою і для виставкової діяльності.

До потенційних переваг використання Інтернету у виставковій діяльності можна віднести такі :

- Широкі можливості використання реклами (банери, різноманітні розсилки, регістрації у пошукових каталогах, рекламні сторінки у виставкових каталогах та ін.).
- Постійний інформаційний обмін між фірмами-учасниками та організаторами, що дає змогу бути постійно в курсі останніх подій та змін.
- Пошук необхідної інформації, партнерів, клієнтів.
- Економія на рекламних матеріалах (використання Інтернет реклами, яка часто є набагато ефективнішою, ніж візуальна реклама), а також економія на послугах комунікації з партнерами/учасниками.

Отже, стаття присвячена створенню логічної структури простору даних у виставковій діяльності як альтернативи забезпечення єдиного засобу зберігання та опрацювання даних.

1. Аналіз літературних досліджень та постановка задачі

Основні ідеї сучасних інформаційних технологій ґрунтуються на концепції баз даних (БД) та сховищ даних. Відповідно до цієї концепції основою інформаційної технології є дані, організовані в БД, що адекватно відображають реалії дійсності у тій або іншій предметній галузі, які забезпечують користувача актуальною інформацією.

Під сховищем даних розуміють особливу базу даних, котра призначена для зберігання в погодженому вигляді історичної інформації, що надходить з різних оперативних систем та зовнішніх джерел. В основу концепції сховища даних покладено розподіл інформації, що використовують в системах оперативної обробки даних (OLTP) і в системах підтримки прийняття рішень.

Існує певний перелік вимог до даних для їх організації та правильної роботи з ними. Серед основних вимог розглянемо взаємозв'язаність даних, яка полягає в тому, що доступ до певної групи даних якогось застосування загалом полегшує доступ до інших груп даних цього самого застосування. В умовах орієнтації БД на велику кількість застосувань виникає необхідність у підтримці значної кількості різноманітних зв'язків між даними. Саме у розумінні тісного логічного зв'язку використовується концепція зберігання даних разом. Вимога мінімізації надлишковості полягає у мінімальній кількості копій для одних і тих самих даних з урахуванням орієнтації на кілька застосувань. Ці надлишкові копії використовуються для підтримки зв'язків між даними. Зайва надлишковість має кілька недоліків. По-перше, зберігання кількох копій веде до додаткових витрат пам'яті. По-друге, доводиться виконувати численні операції оновлення для кількох надлишкових копій. Крім того, оскільки різні копії даних можуть відповідати різним стадіям оновлення, то інформація, що зберігається в системі певний час, може стати суперечливою. Під цілісністю БД розуміють несуперечливість між собою даних, що в ній зберігаються.

Протягом десятиліть співтовариство управління даними концентрувалося на розробленні систем управління реляційними базами даних, досягнувши вражаючих результатів. Проте останніми роками потреби в даних, що швидко розширюються, привели до виникнення галузі, в якій ведуться дослідження, які є цікавими і продуктивними, але не об'єднані якою-небудь основною метою або узгодженою програмою роботи. Сьогодні найгостріші проблеми управління інформацією "виростають" з організацій (наприклад, комерційних підприємств, урядових агентств, бібліотек), що покладаються на велике число різнотипних, взаємозв'язаних джерел даних, але що не мають можливості управляти цими просторами даних зручним, інтегрованим і обгрунтованим способом. У цій статті як нова програма робіт в області управління даними пропонуються простори даних і підтримувальні системи.

За сучасними сценаріями управління даними частіше всього дані не знаходяться під управлінням традиційної реляційної СУБД або якої-небудь іншої моделі даних або системи. Натомість розробники часто стикаються з набором слабо зв'язаних джерел даних і тому кожного разу розв'язують низькорівневі задачі управління даними, що повторюються, в різнорідних колекціях. До цих завдань належать забезпечення можливостей пошуку і запиту даних; дотримання правил, обмежень цілісності погодження іменування і т.д.; відстежування походження даних; забезпечення доступності, відновлення і контролю доступу; керований розвиток даних і метаданих.

Варто ввести поняття простору даних як нової абстракції управління даними в таких сценаріях. Як ключову програму робіт у галузі управління даними ми пропонуємо проектування і розроблення платформ підтримки просторів даних (DataSpace Support Platforms, DSSP). Інакше кажучи, DSSP забезпечує набір взаємозв'язаних послуг і гарантує розробникам можливість концентруватися на специфічних проблемах їх додатків, а не на завданнях, що повторюються, виникають за потреби узгодженої та ефективної роботи з взаємопов'язаними, але роздільно керованими даними [4, 5].

2. ОСНОВНИЙ МАТЕРІАЛ

2.1. Формальний опис простору даних

Ми пропонуємо нову систему, яка може обробити повний особистий простір даних (Dataspace) окремого користувача чи підприємства. Розглянемо використання просторів даних у виставковій діяльності. Як і особистий простір даних, так і простір даних підприємства виставкової діяльності (Dataspace) містить всі дані, що стосуються користувача (групи користувачів) на всіх його дисках і на віддалених серверах – як, наприклад, мережеві драйвери, електронна пошта, веб-сервери і таке інше. Ці дані представлені різнорідним змішуванням файлів, електронної пошти, закладок, музики, зображень, календарних даних, особистих інформаційних потоків тощо.

Отже, *простір даних* **DS** – це множина даних, поданих у різних моделях (баз даних **DB**, сховищ даних **DW**, статичних веб сторінок **Wb**, неструктурованих даних **Nd**, графічних та мультимедійних даних **Gr**), локальних сховищ та індексів (**ODW**), а також засобів інтеграції (**Int**), пошуку (**Se**) та опрацювання інформації (**Wo**), об'єднаних середовищем управління моделями (**EM**).

DS=<**DB**, **DW**, **ODW**, **Wb**, **Nd, Gr**, **Int**, **Se**, **Wo, EM**>

Сучасні інструменти, наприклад, пошук на робочому столі і настільні операційні системи (зокрема Vista) не здатні вирішувати як проблему фізичної особистої інформаційної незалежності (де мої дані), так і формат та незалежність моделі даних (як вони збережені і які програмні ресурси мені доведеться використовувати для того, щоб звернутися до цих даних). Наша робота ґрунтується на використанні єдиних систем для управління особистою інформацією [1], і роботі [2], яка описує простори даних (dataspaces) як нову абстракцію для інформаційного управління.

Простори даних недавно були ідентифіковані як новий порядок для інформаційного управління [1–3]. Отже, система управління простором даних є новим видом архітектури інформаційного управління, яка дає змогу управляти всіма даними, що мають відношення до певної організації, виду діяльності, завдання або людини. У чіткому контрасті до існуючих архітектур інформаційної інтеграції, система управління простором даних є підходом співіснування даних: він не вимагає семантичної інтеграції, поки послуги на дані забезпечені. Для прикладу, пошук за ключовим словом має підтримуватись у будь-який час для всіх даних (схема пізніше або схема ніколи). Існуючий простір даних у такому випадку може бути поступово розширений, визначаючи

семантичні зв'язки між різними компонентами простору даних. Як конкретний приклад використання просторів даних можна розглянути особистий простір даних, тобто всі електронні пристрої, які належать єдиній людині. А якщо говорити про виставкову діяльність, то такими елементами, що будуть входити у простір даних, є інформація про виставкове обладнання, учасників виставок, продукцію, що виставляється на виставці, відвідувачів виставок тощо.

2.2. Концептуальна модель простору даних виставкової діяльності

Розглянемо, наприклад, підприємство, яке працює в сфері виставкової діяльності та займається проведенням та організацією виставок та ярмарків як міжобласного так і міжнародного характеру. Однією важливою складовою проведення тієї чи іншої виставки є її прогноз на майбутнє, тобто чи буде вона прибутковою, чи збитковою. Цим займається певна група людей, яка слідкує за різноманітними показниками виставкової діяльності, зокрема і порівняно з іноземними фірмами.

Спостереження за відвідувачами і учасниками виставок, моделювання і планування виставкових робіт забезпечує вхідні дані для програм і аналізів, які генерують широкий діапазон продуктів даних, які використовуються для майбутнього прогнозування виставкової тематики, наприклад, на наступний рік.

Такі дослідницькі групи володіють величезною кількістю корисної інформації, яка може бути використана і іншими подібними групами, тобто, дослідницькі групи кількох виставкових організацій можуть плідно обмінюватись інформацією. Такі групи об'єднуються для створення виставкових просторів даних регіонального або національного масштабу. Їм треба буде якомога простіше імпортувати свої дані в стандартних форматах і з глибиною деталізації (частина файла або кілька файлів).

Для швидкого пошуку в таких колекціях даних можуть бути корисними локальні копії або додаткові індекси.

Цей сценарій ілюструє кілька вимог простору даних :

- каталог простору даних;
- підтримку аналізу походження даних
- створення колекцій та індексів

Простір даних повинен містити всю інформацію, необхідну для конкретної організації, незважаючи на формат і місцерозташування цієї інформації, а також моделювати розвинений набір зв'язків між даними. Отже, ми моделюємо простір даних як набір учасників і зв'язків.

Учасниками виставкового простору даних є індивідуальні джерела даних: вони можуть бути реляційними базами даних, репозиторіями XML, текстовими базами даних, Web-сервісами і пакетами програмного забезпечення. Вони можуть зберігатися або бути потоками даних (локально керованими системами потоків даних), або навіть сенсорними установками.

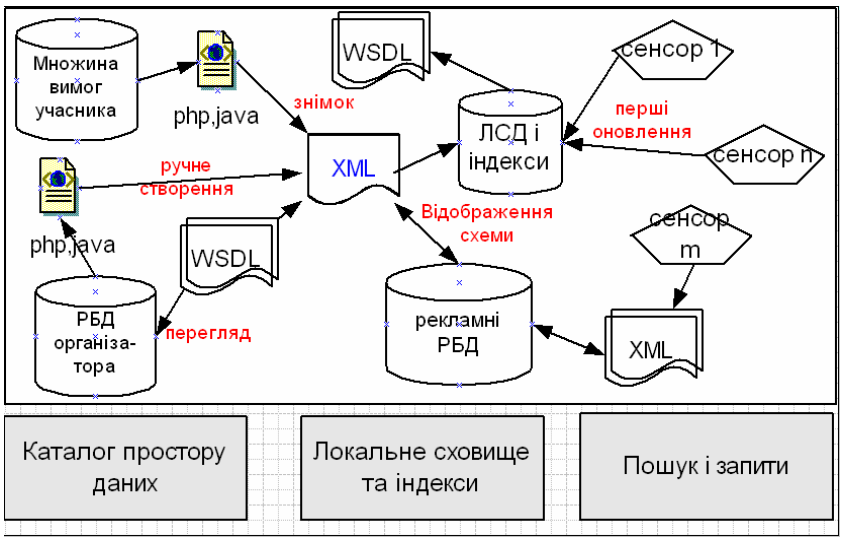

Рис. 1. Приклад простору даних виставкової організації і компоненти системи простору даних

Деякі учасники можуть підтримувати виразні мови запитів, а інші – бути неінтелектуальними і такими, що підтримують лише обмежені інтерфейси для формулювання запитів (наприклад, структуровані файли, Web-сервіси або інші програмні пакети). Учасники можуть бути структурованими (наприклад, реляційними базами даних), напівструктурованими (XML, колекції коду) або повністю неструктурованими. Деякі джерела підтримуватимуть традиційні операції оновлення, інші – допускають тільки додавання (з метою архівації), а треті можуть бути повністю немінливими.

Двома основними службами, які підтримуватимуться в DSSP, є пошук і запит даних, тоді як СУБД відрізняються покращеною підтримкою запитів, а пошук є основним механізмом роботи кінцевих користувачів з великими колекціями незнайомих даних. Пошук менш вимогливий, ніж запит даних, оскільки він заснований на схожості, наданні кінцевим користувачам ранжованих результатів і підтримці інтерактивного вдосконалення, так що користувачі можуть досліджувати набір даних і покращувати свої результати. DSSP повинні допомагати користувачам задавати пошуковий запит та ітераційно його удосконалювати, якщо це доречно, відносно виду запиту в стилі бази даних. Ключовий принцип просторів даних полягає в тому, що пошук повинен бути застосовний до всього вмісту простору даних, незалежно від форматів даних.

Інші ключові сервіси DSSP містять моніторинг, виявлення подій і підтримку складних потоків робіт. Наприклад, ми хочемо визначити виставкову площу під час надходження нової частини даних і поширити результати цього обчислення на набір приймальних джерел даних. Аналогічно, в DSSP повинні підтримуватися різні форми аналізу даних.

Ключові проблеми, що виникають при створенні компоненту DSSP локального зберігання і індексації, пов’язані з неоднорідністю індексу. Індекс повинен одноманітно індексувати всі можливі елементи даних: чи є вони словами, що зустрічаються в тексті, значеннями, що зустрічаються в базі даних, або елементом схеми одного з джерел. Крім того, в індексі повинна передбачатися можливість наявності декількох способів посилання на один і той самий об’єкт реального світу. Поки що дослідження у галузі узгодження посилань фокусуються на визначенні ситуацій, коли декілька посилань належать до одного і того самого об’єкта).

Складно буде підтримувати індекс в актуальному стані, особливо для учасників, що не мають механізмів сповіщення про оновлення.

Менеджер

Наша система повинна забезпечити інтерфейс між користувачем і учасниками простору даних. Менеджер – це центральний компонент, який відповідальний за взаємодію з користувачем. Він проводить аутентифікацію користувача і займається призначенням прав.

Локальна Пам’ять та Індекс

Цей компонент управляє кешуванням пошуку і представляє результати так, що відповіді на ці запити можуть бути представлені без доступу до фактичних даних. Це створює ефективні асоціації на запити між об’єктами даних різних учасників і покращує доступ до джерел даних з обмеженим доступом. Індекс має бути якнайкраще адаптованим до різнорідних середовищ.

Це робиться так, що будь-яка поява елементу в просторі даних фіксується, і повертається його розташування (місцезнаходження), у якому він з’являється, та роль кожної події (рядок в текстовому файлі, елемент в шляху до файла, або тег у файлі XML). Так інформація розноситься всім залученим учасникам. Інший важливий аспект – стійка обробка множинних посилань на об’єкти реального світу (різні способи послатися на компанію або людину). Кешування збільшує здатність даних, які зберігаються в учасників простору, які можуть і не бути надійними, і скорочують завантаження запиту учасників, які не можуть дозволити у цьому випадку зовнішні запити.

Архітектура системи управління простором даних виставкової діяльності показана на рис. 2. Він містить трьох учасників, враховуючи різні моделі даних: RDB, XML і файловий архів. Спочатку кожен учасник реєструється в Каталозі, і він стає доступним як джерело даних. Відповідальний за реєстрацію – Менеджер – так само, як і для створення зв’язків між ними, для дозволу користувачу удосконалити ці взаємозв’язки, для аутентифікації користувача і передачі прав доступу. DSMS не приймає на себе повний елемент управління над даними. Натомість це дає змогу окремим системам управляти даними, але забезпечує безліч послуг, щоб регулювати деякі відсутні особливості управління. Існує три внутрішні контейнери даних, а саме Архів Метаданих (MDR – Metadata Repository), Пам’ять

Дублювання (RS – Replication Storage), Локальна Пам'ять та Індексація (LSI – Local Storage and Indexing), які дають користувачам змогу перенести дані між компонентами і локальними машинами. Витягнута інформація (за допомогою Екстрактора Даних) є локально збережена та індексована. Процесор Запиту стикається з проблемою глобального запиту на множинних базах даних на множинних сайтах. Інтерпретатор Запиту транслює запит в мови, підтримувані учасниками.

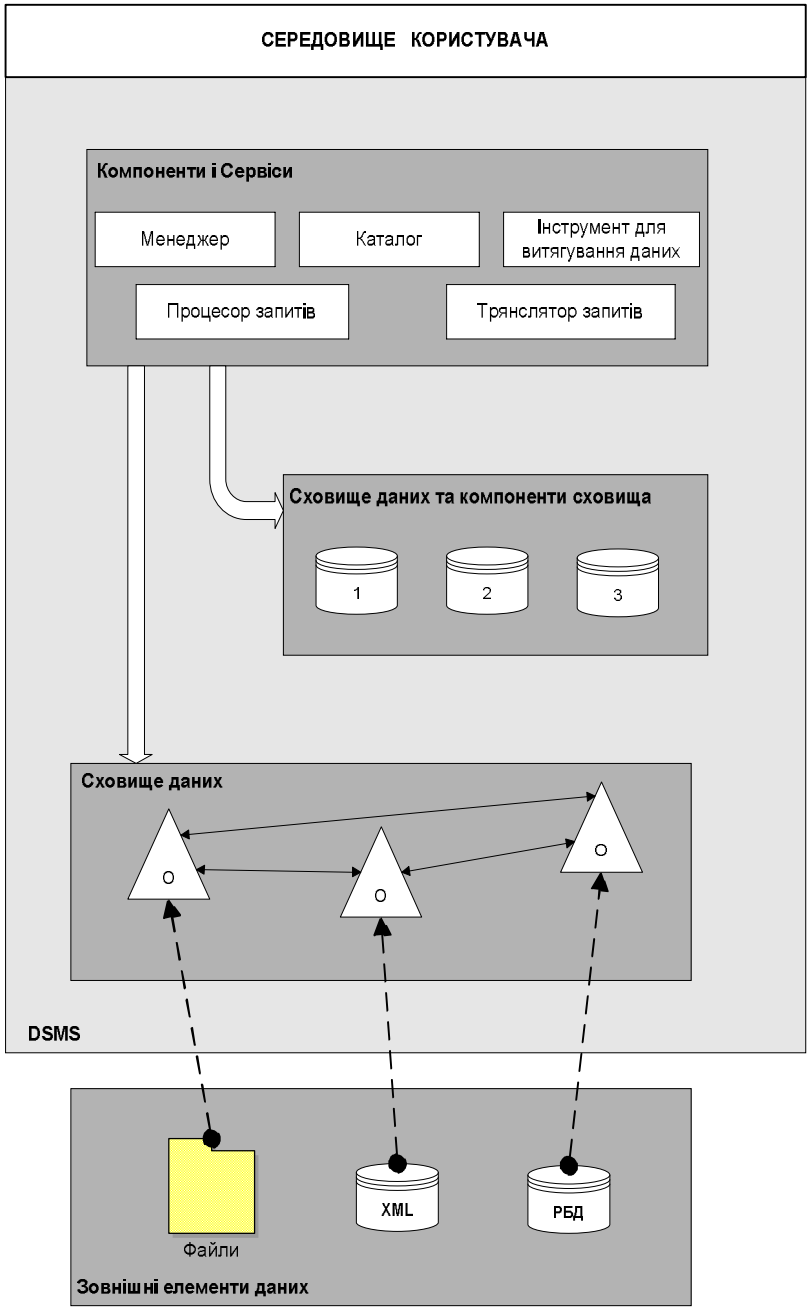

Рис. 2. Архітектура Системи управління простором даних

Персональне Інформаційне Управління (PIM) PIM повинне забезпечити легкий доступ до всієї персональної інформації, що збережена на робочому столі, з можливим під'єднанням до мобільних пристроїв, особистої інформації у Web, або навіть інформації, до якої звертались протягом всього життя особи. Ідея полягає в розширенні пошуку на робочому столі (зараз він обмежений запитами ключового слова), використовуючи асоціації між незрівнянними елементами на робочому столі, наприклад "Рахують складену рівновагу мого банківського рахунку". Додатково ми хотіли б зробити запит про джерела, наприклад "Пошук всіх експериментів, запущених студентом X". Наступні принципи

просторів даних dataspaces тут: інструмент PIM повинен надати доступ до всіх даних, щоб гарантувати інтеграцію множинних джерел даних і повернути найкращий результат. Прикладами реалізації системи PIM є iMeMex і SEMEX.

Реляційна база даних

Нижче продемонстрований приклад реляційної бази даних виставкового підприємства (реалізований в СУБД MS Access 2005):

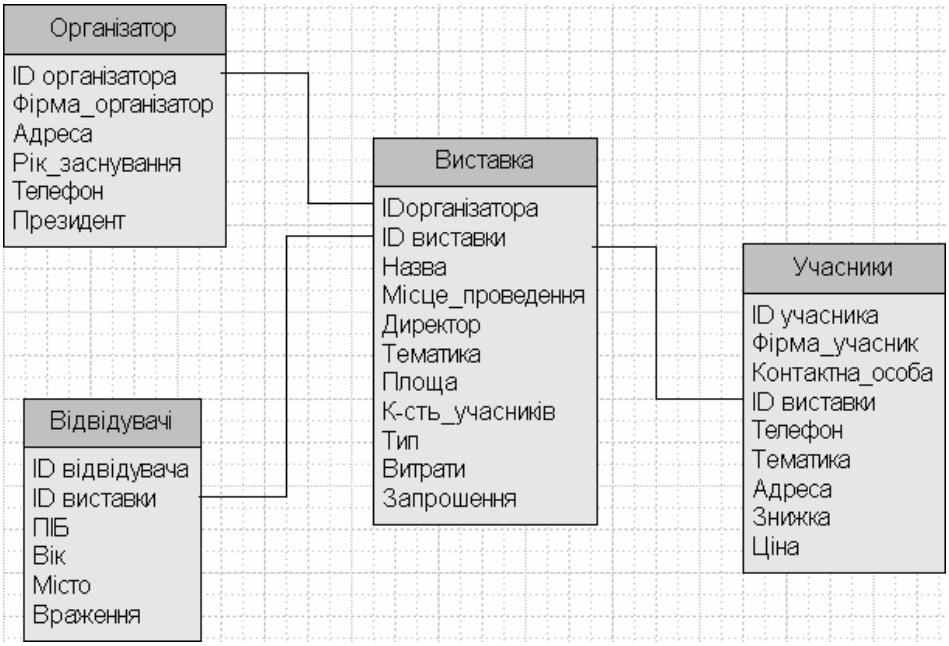

Рис. 3. Можлива логічна модель даних виставкової організації

DSSP забезпечує декілька взаємозв'язаних служб над простором даних, деякі з яких є узагальненням компонентів, підтримуваних в традиційній СУБД. На відміну від СУБД, в DSSP не передбачається наявності повного контролю над даними в просторі даних. Натомість, DSSP дає змогу управляти даними системам-учасникам, але забезпечує новий набір служб, враховуючи їхню потребу в автономності.

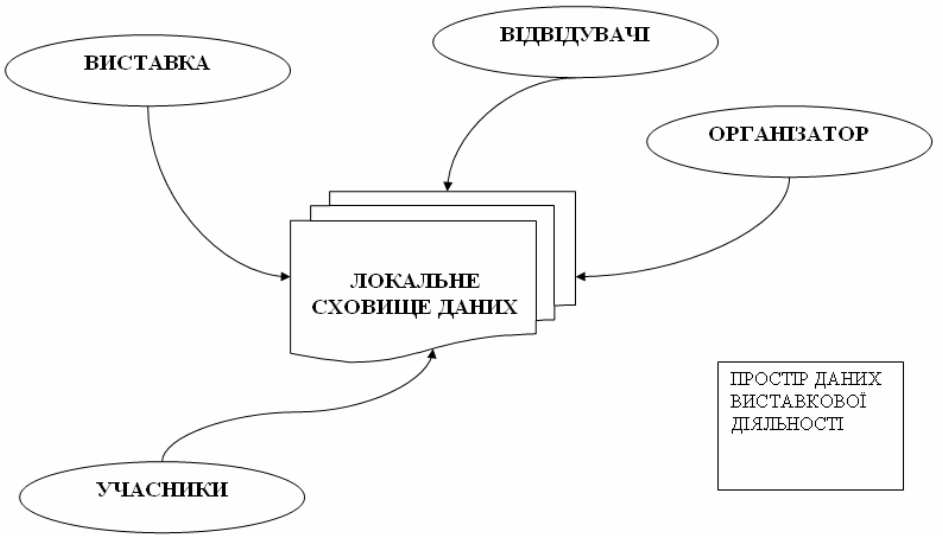

Рис. 4. Локальне сховище даних виставкової організації

Схематично локальне сховище даних організації, яка займається виставковою діяльністю, можна зобразити за допомогою структурного рисунка (рис. 5):

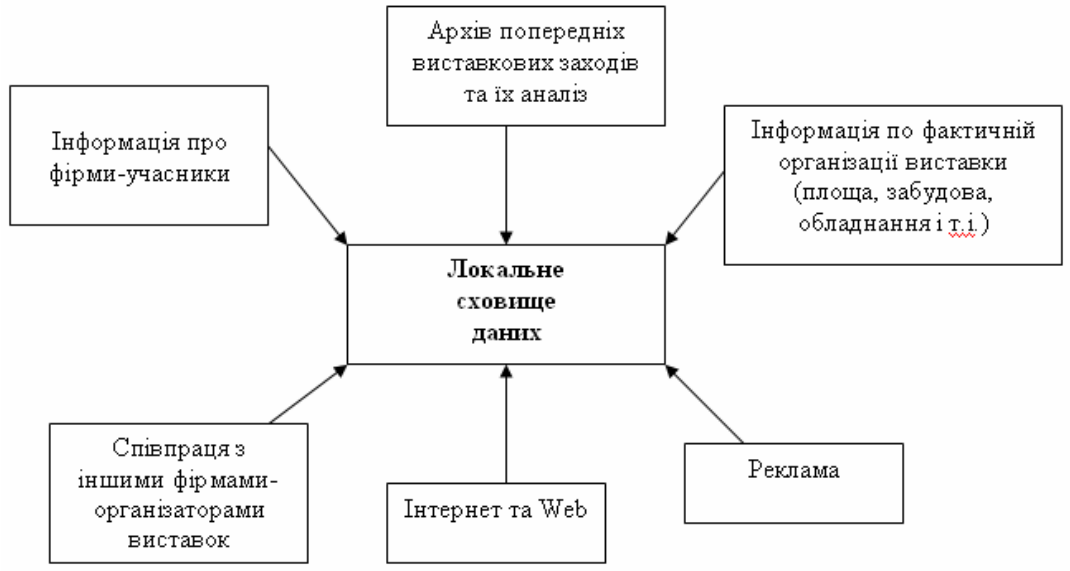

Рис. 5. Структурна схема локального сховища даних

Висновки

Описано простір даних фірми, яка займається виставковою діяльністю, а саме організацією та проведенням виставок, що забезпечує взаємодію між джерелами інформації, поданої за допомогою різних моделей даних з різними способами подання та опрацювання.

Наукова новизна. Новизна полягає у впровадженні формального опису простору даних та окресленні його основних задач.

Практична цінність полягає у побудові простору даних виставкової діяльності, виділенні основних об'єктів та учасників.

Подальші дослідження стосуватимуться формалізації методів інтеграції даних та пошуку неструктурованих, напівструктурованих та строго структурованих даних.

1. Garretts Summary of Principles of Dataspace Systems / [Електронний ресурс] / Arizona State University – Режим доступу: http://aravaipa.eas.asu.edu/wiki/index.php/Garretts_Summary_of_Principles_of_Dataspace_Systems#Overview. 2. Kossmann D. Personal Data Space / [Електронний ресурс]/ Dittrich J.-P. – Depeartment Informatik. –Режим доступу: http://www.inf.ethz.ch/news/focus/res_focus/feb_2006/index_DEProf. Donald Kossmann, Jens-Peter Dittrich. Personal Data Spaces. 3. Kingsley Idehen. Semantic Web Data Spaces. Web Data Spaces, 2007. 4. Огляд технологій інтеграції інформаційних систем / [Електронний ресурс] /Microsoft.com. – 2006. – Режим доступу: http://www.microsoft.com/Ukraine/Government/Analytics/IntegrationTechnologies/Overview.mspx. 5. Кузнецов С. От баз данных к пространствам данных: новая абстракция управления информацией : Информационно-аналитические материалы / [Електронний ресурс] / Кузнецов С. Д. – Центр информационных технологий, 2001. – Режим доступу : http://www.citforum.ru/database/articles/from_db_to_ds. 6. Dan E. Linstedt. Data Vault Overview: The Next Evolution of Data Modeling [Електронний ресурс] / Dan E. Linstedt. – July 1, 2002. — Режим доступу: http://www.tdan.com/i021hy01.htm. 7. Madden S. TinyDB: An Acquisitional Query Processing System for Sensor Networks / [Електронний ресурс] / Madden S., Franklin M. J., Hellerstein J. M., Hong W.. ACM TODS, 30, No. 1. – March 2005. — Режим доступу: http://db.csail.mit.edu/madden/html/tinydb_tods_final.pdf. 8. Кэтэрин Дрюэк (Katherine Drewek). "Хранилища данных: сходство и различия подходов Билла Инмона и Ральфа Кимболла",[Електронний ресурс] / 2005, — Режим доступу: http://www.b-eye-network.com/view/743