

УДК 004.75

П.П. Маслянюк, П.М. Лісов

## ДОСЛІДЖЕННЯ ЗАСОБІВ І РОЗРОБКА ТЕХНОЛОГІЇ ПРОДУКУВАННЯ ІНФОРМАЦІЙНИХ РЕСУРСІВ

### Вступ

На сьогодні одним з ефективних засобів підвищення конкурентоздатності організаційної системи (Орг.С) є її інформатизація. Як зазначалося в [1–3], під визначенням “інформатизація організаційних систем” розуміємо необхідну і достатню множину правових, організаційних, економічних, наукових та науково-технічних рішень і процесів, спрямованих на створення інформаційно-комунікаційних систем з метою задоволення інформаційних потреб, забезпечення і автоматизації бізнес-процесів, підтримки прийняття рішень та підвищення ефективності керування організаційною системою із застосуванням інформаційно-комунікаційних технологій.

Інформатизація передбачає створення інформаційно-комунікаційної системи (ІКС) [4]. Однією з компонентів цієї системи є інформаційний ресурс, який забезпечує зберігання та обробку даних, інформації і знань як одну з основних цінностей Орг.С [5]. Для того щоб створити ресурс, який буде максимально ефективно задовольняти потреби організації, необхідно провести ґрунтовний аналіз як згаданих потреб, так і характеристик та особливостей інформації.

Конкурентоспроможність будь-якої організаційної структури, підприємства чи фірми прямо залежить від кількості і якості інформації, якою вони володіють, а також від наявності засобів оперування такою інформацією. Тому проблема створення інформаційних ресурсів є важливою складовою для забезпечення конкурентоспроможності Орг.С.

### Постановка задачі

Метою статті є дослідження засобів і розробка технології продукування інформаційних ресурсів. У зв'язку з цим необхідно запропонувати підхід, який дозволить обґрунтовано вибрати засоби та технології для створення інформаційних ресурсів. Для цього треба проаналізувати дані, інформацію і знання, які можуть збе-

рігатись в інформаційних ресурсах Орг.С. На основі прикладного системного аналізу і компонентного процесу розробки слід розробити підхід до проектування і створення ресурсу як системи компонент, певним чином пов'язаних між собою та з іншими компонентами ІКС. Мають бути досліджені інформаційні ресурси, їх класифікація, основні типи.

Результати аналізу засобів продукування інформаційних ресурсів є основою для формулювання рекомендацій по застосуванню тих чи інших засобів для продукування окремих компонентів ресурсу ІКС.

Результати статті спрямовані на застосування при розробці проектів інформатизації організаційних систем, зокрема для системної інженерії інформаційних ресурсів.

### Інформаційний ресурс

Однозначного і завжди коректного визначення понять “дані”, “інформація”, “знання”, на жаль, немає. Сам термін “інформація” є порівняно недавнім: до другої половини ХХ ст. в російській і українській мовах це поняття не вживалось [6]. З латинської *informatio* перекладається як відомості або роз'яснення. Можна навести багато визначень інформації – від філософського – “відображене різноманіття, яке виникає в результаті взаємодії між об'єктами” [7] до визначення Лебедева: “послідовність нулів та одиниць”. Однак перше з них не є конструктивним, а друге не зовсім відповідає сучасному розумінню інформації.

Згідно із законодавством України, інформація – це відомості, подані у вигляді сигналів, знаків, звуків, рухомих або нерухомих зображень тощо [8]. Законодавство України визначає дані як інформацію у формі, придатній для автоматизованої обробки її засобами обчислювальної техніки [8].

За європейськими стандартами, знання – це комбінація даних та інформації, до яких додається точка зору, навички і досвід експерта, що дає вагомий результат, який можна використати для прийняття рішень. Знання може бути вичерпним та/або вузьким, індивідуальним та/або колективним. На жаль, українське законодавство взагалі не визначає поняття “знання”.

Згідно з іншими джерелами, “дані – це інформація до того, як їй надається контекст, структура та значення” – “data is information before it has been given any context, structure and meaning” [9].

На нашу думку, найбільш вдалим є визначення, за яким дані – це результат простого збору визначених фактів; інформацією вони стають лише при зв'язуванні у щось корисне комбінації: хто, що, де і як. У свою чергу, знання – це розуміння того, як і чому щось відбувається [10]. Саме такі значення понять “дані”, “інформація” і “знання” будуть використовуватись у нашому дослідженні.

Дані, інформація і знання – сутності, які можуть відображати об'єкти, процеси та явища навколишнього середовища. Вони є абстрактними об'єктами [6]. Засобом їх матеріалізації є інформаційний ресурс. Згідно із законодавством України, інформаційний ресурс – це сукупність документів в інформаційних системах (бібліотеках, архівах, банках даних тощо) [11].

Документ – це упорядкована сукупність даних, інформації і знань, яка надає можливість доступу, передачі, обробки тощо. Прикладом його може бути паперовий документ, фільм, комп'ютерний файл [6].

Середовищем зберігання документів є інформаційно-комунікаційна система, яка забезпечує доступ, обмін інформацією та її обробку. Така система не обов'язково повинна бути комп'ютеризована. Прикладами комп'ютеризованих інформаційно-комунікаційних систем можуть бути операційне середовище, база даних, інформаційне сховище та інформаційний колектор.

Продуктування інформаційних ресурсів є одним з основних завдань інформаційного суспільства [12]. Під цим поняттям розуміємо створення нових видів інформаційних ресурсів на основі існуючих даних, інформації і знань.

Класифікація інформаційних ресурсів дозволяє розбити ресурс на компоненти, визначити характеристики та вибрати оптимальні засоби для їх реалізації. Для опису вимог до інформаційного ресурсу, що розробляється, використовують такі його характеристики [13]:

- характеристики продуктивності:
  - пропускна спроможність;
  - час реакції;
  - час затримки;
- характеристики надійності;
- характеристики масштабованості;
- характеристики розширюваності;
- характеристики прозорості;
- повна вартість володіння.

Повна вартість володіння інформаційним ресурсом визначається не тільки (і не стільки) вартістю використаних при його створенні апа-

ратних і програмних засобів, а й вартістю інформації, яка в нього закладена.

Класифікація інформаційних ресурсів дає можливість обґрунтовано визначити також необхідний тип інформаційного ресурсу, вибрати інструменти і технології для його продукування.

Інформаційні ресурси можна класифікувати за такими ознаками [14]:

- приналежність ресурсу до певної організаційно-технологічної системи (наприклад, бібліотечної мережі, ЗМІ, корпоративної системи);
  - спосіб виділення об'єктів обліку (творів, документів, видань, баз даних, інтернет-сторінок, сайтів тощо);
  - призначення ресурсу (масової інформації, освіти, бізнесу, особистої переписки тощо);
  - зміст ресурсу:
    - тематичний;
    - об'єктний;
    - функціональний;
  - видовий склад ресурсу (види документів);
  - джерело інформації:
    - національне або закордонне;
    - офіційне або неофіційне;
    - інші джерела;
  - правовий статус ресурсу (публічні документи, об'єкти інтелектуальної власності, спам, таємні документи тощо);
  - структурний тип ресурсу, що вміщує в собі:
    - можливість відділення даних від програм та представлення;
    - формати;
    - кодування;
    - інше;
  - відкритість ресурсу (відкритий або з обмеженим доступом);
  - рівень структурованості:
    - структуровані;
    - неструктуровані;
  - спосіб поширення і носій;
  - мова ресурсу.
- Комп'ютеризовані інформаційні ресурси за змішаним критерієм щодо надання сервісів і можливостями оперувати даними, інформацією та знаннями, можна розділити на такі основні типи [13]:
- файлові системи;
  - бази даних;
  - інформаційні сховища;
  - інформаційні колектори;
  - веб-ресурси.

**Файлові системи** (ФС) є найпростішим і найбільш поширеним типом інформаційних ресурсів. Вони зберігають дані, інформацію і знання довільного типу та довільної структури. ФС, з одного боку, є персональним інформаційним ресурсом, а з іншого – базою для інших типів ресурсів.

Розвиток ФС розпочався із створення носіїв даних досить великого розміру та виникнення задач, для яких вже не можна було зберігати на носії дані одного типу у визначеному форматі.

Можна виділити такі типи ФС [15]:

- *дисккові* – зорієнтовані на зберігання розрізаних даних на одному жорсткому диску персонального комп'ютера користувача; здебільшого вони використовуються на ПК для розв'язання персональних задач користувача;

- *кластерні* – зорієнтовані на зберігання даних на масиві дисків, що забезпечує вищу швидкість, надійність і ємність. Вони, як правило, використовуються в серверах, зараз їх все частіше застосовують в ПК (RAID-масиви);

- *мережні* – дозволяють пов'язати в єдину ФС носії на різному апаратному забезпеченні, поєднані в межах мережі. Розвиток таких систем дає можливість створити інформаційний ресурс підприємства із забезпеченням максимальної надійності і швидкості роботи при уникненні дублювання;

- *спеціального призначення* – до цієї групи належать ФС, оптимізовані для роботи на різноманітних носіях (компакт-дисках), системи доступу до ftp, системи файлових пристроїв та ін.

Ми розглядаємо два основні класи задач, які розв'язуються за допомогою ФС: організацію розміщення даних на фізичному носії та організацію файлового простору користувача (ієрархії файлів та каталогів).

Фізична організація зберігання даних на диску передбачає, що для користувача (програми) створюється інтерфейс за відповідними правилами, тобто певний образ організації файлів і каталогів. Реальне місце кожного файла встановлюється ФС за певним алгоритмом, який має забезпечити швидкий доступ до файлів, пошук, запис, операції переміщення/копіювання тощо. Це завдання передусім пов'язане із фізичною організацією пристроїв для збереження даних.

Другим аспектом розробки ФС є організація ієрархії файлів і каталогів. Вона має забезпечити структурованість інформації, швидкість і зручність її обробки та підтримку стан-

дартних рішень. Як правило, ієрархію файлів кожний користувач визначає самостійно, щоправда, інколи існують рішення, зорієнтовані на певну визначену ієрархію. Прикладом цього може бути організація ФС для операційної системи (ОС) Windows. Користувач може створювати ієрархію для своїх даних, проте для ОС використовується ієрархія, створена розробником. Аналогічно, якщо в межах організаційної структури використовуються певні загальні програмні рішення, то може виникнути необхідність у розробці ієрархії ФС. Така розробка має проводитись на етапі проектування, а надалі за допомогою організаційних засобів повинна контролюватись підтримка встановленої ієрархії.

Останнім часом на файлові системи покладаються нові задачі: контроль доступу до даних, підтримка політик безпеки, архівація та ін. Це викликано тим, що ФС у сучасних системах вже не є лише основою операційної системи, а являє собою окремий інформаційний ресурс, доступ до якого може проводитись і "в обхід" ОС. Тому функції, які раніше покладалась на ОС, тепер мають бути перекладені на саму ФС.

**База даних** (БД) – це інформаційний ресурс, призначений для зберігання структурованих даних. Класифікуються БД за структурою інформації, що зберігається:

- *каталоги* – найпростіший вид БД, в якому використовується підхід, аналогічний до бібліотечних каталогів. При цьому вся інформація зберігається в одній таблиці з визначеним набором полів. Такі БД були історично першими і почали розвиватись у 60-х роках ХХ ст.;

- *ієрархічні*; у таких БД дані організовані в деревоподібну структуру, в якій кожен запис є вузлом. Ця організація БД забезпечує швидкий пошук даних (за одним критерієм);

- *мережні*; такі БД зберігають дані із зв'язками з іншими даними. Зв'язки організуються як вказівники (pointer) на інші записи. Як правило, для зберігання записів створюється схема, схожа на ієрархічну;

- *реляційні* – найпопулярніші сучасні БД, які базуються на реляційній моделі, запропонованій у 1969 р. Е. Коддом [16]. В них дані подані у вигляді таблиць, кожен рядок яких є одним записом, а кожний стовпчик – визначеним атрибутом. Робота з БД відбувається за допомогою запитів, результатом яких є таблиці;

- *пост-реляційні*; такі бази використовують деякі принципи реляційних БД, але не обмежуються реляційною моделлю. Часто вони використовують також принципи ієрархічних та мережних БД;

• *об'єктні*; при врахуванні розвитку об'єктно-орієнтованого програмування виробляються спроби створити схему БД, яка б дозволила працювати з об'єктами. У такому випадку БД і програма, яка її використовує, мають однакові назви, що дає можливість спростити взаємодію і уникнути проблем із перетвореннями даних.

Слід зазначити, що при розробці сучасних систем інколи доречно використовувати БД із змішаною структурою [17]:

- за характером інформації, яка зберігається:
  - фактографічні (картотеки);
  - документальні (архіви);
- за способом зберігання даних:
  - централізовані (зберігаються на одному комп'ютері);
  - розподілені (використовуються в локальних і глобальних комп'ютерних мережах).

**Інформаційні сховища** – це ресурси, призначені для аналітичної обробки інформації. За визначенням Білла Інмона, сховище даних є предметно-орієнтованим, хронологічним, стабільним набором даних, призначеним для підтримки прийняття управлінських рішень [18]. З іншого боку, інформаційне сховище являє собою ієрархічно організовану сукупність БД, призначених для збереження і обробки архівних даних.

Бази даних дають можливість ефективно працювати з оперативними даними, але для аналітичних запитів час обробки може бути значним, оскільки доводиться опрацьовувати великі об'єми інформації. Натомість, в інформаційне сховище додаються оброблені певним чином дані, тобто первинна аналітична обробка проводиться постійно і поступово. Таким чином, отримується багатовимірна структура, яка не містить у собі оперативних даних, але містить матеріал для аналітичної роботи. На запит експерта сховище дозволяє швидко побудувати так званий зріз, який має необхідну інформацію.

На сьогодні для більшості Орг.С вже недостатньо обробки лише даних та інформації – необхідною є робота із знаннями. Для зберігання і обробки знань використовуються *інформаційні колектори* [19].

Першими інформаційними колекторами були бібліотечні системи. Вони зберігали тексти документів у сховищах бібліотек [12]. Документи були класифіковані за стандартними бібліотечними класифікаторами. В теперішніх інформаційних колекторах для пошуку документів використовуються метадані [20], що спри-

яє більш ефективно і швидко знаходити документи.

У сучасних умовах дуже важливим стало завдання інтеграції колекторів. Це пов'язано з характером знань, що зберігаються в колекторах [19]. В більшості колекторів знаходиться саме відкрита інформація. Створення єдиної мережі колекторів [21, 22] сприяє забезпеченню можливості пошуку інформації в різних колекторах без необхідності проводити його в кожному окремо.

На сьогодні інформаційні колектори використовуються для збереження знань у різних формах. Сучасні системи підтримують зберігання не лише текстів, а й відео, аудіо та інших мультимедійних документів.

**Веб-ресурси** являють собою ресурси, які зберігаються на різних компонентах, розподілених у гетерогенній мережі. Прикладом одного з таких ресурсів є мережа Інтернет. Перевагою даного типу ресурсу є надійність та забезпечення зручного доступу, недоліком – складність організації керування і пошуку.

Як видно з аналізу, кожний із розглянутих типів інформаційних ресурсів призначається для вирішення певних завдань. При побудові ІКС Орг.С необхідно вирішувати різні типи завдань, тому в більшості випадків використовується комбінований ресурс, який поєднує в собі систему окремих взаємодіючих ресурсів. Компонентний підхід дозволяє описати такі ресурси і їх взаємодію на рівні інтерфейсів. Типову модель організації ресурсів у складі інформаційного ресурсу ІКС зображено на рис. 1.

На ньому показано інтерфейси взаємодії як із зовнішніми для ресурсу компонентами (порталом, зовнішнім ресурсом), так і між внутрішніми компонентами (ресурсами) в межах комплексного ресурсу.

Корпоративний портал являє собою програмний продукт, який за допомогою браузера забезпечує безпосередній доступ до розподілених інформаційних ресурсів: файлових систем, інтегрованих баз даних, інформаційних сховищ та інформаційних колекторів [1].

Корпоративний портал – це функціонально повна частина ІКС організації, яка надає користувачам єдину точку доступу до внутрішніх та зовнішніх інформаційних ресурсів, необхідних для прийняття обґрунтованих управлінських рішень.

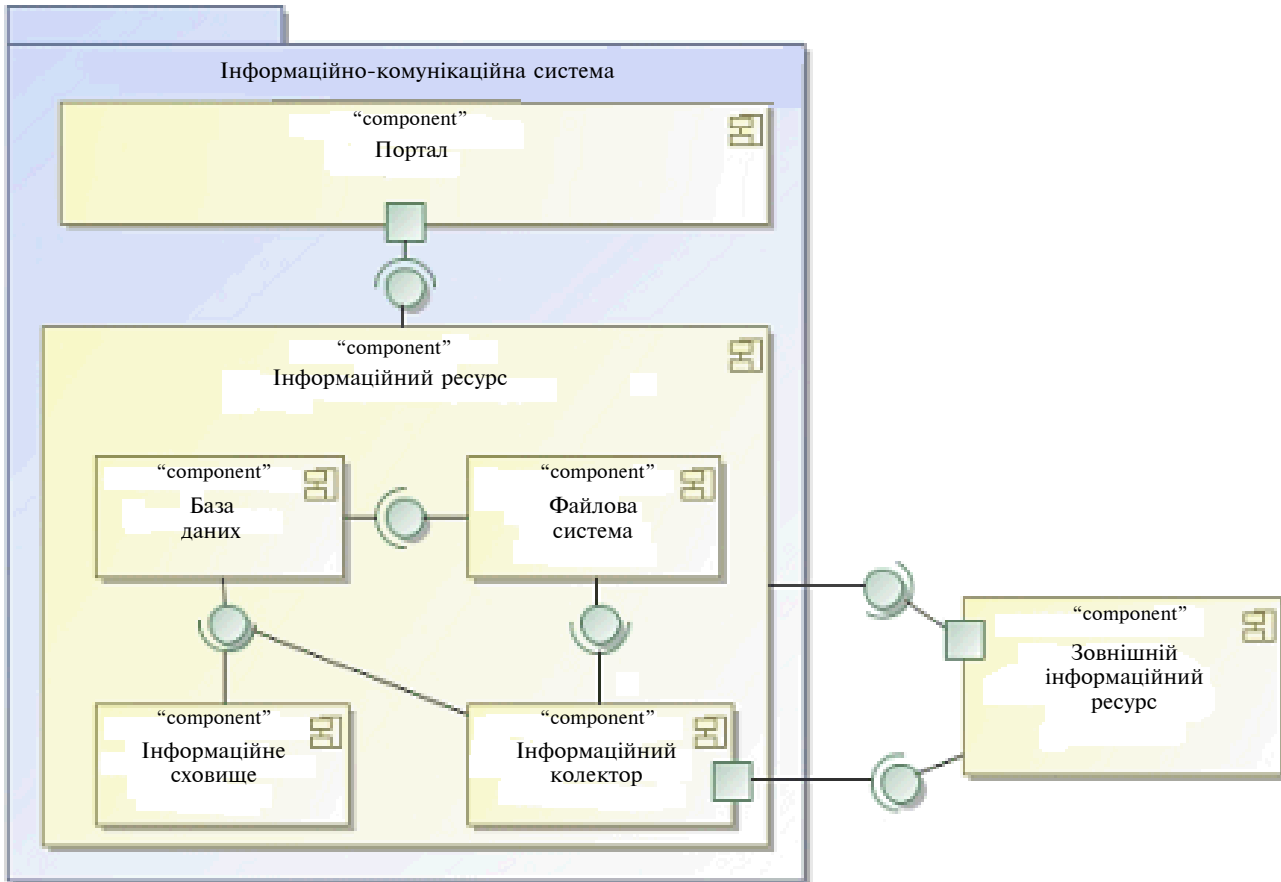


Рис. 1. Модель організації інформаційного ресурсу ІКС. Діаграма компонентів у нотатції UML

### Інтеграція і інтероперабельність інформаційних ресурсів

Розробка ізольованих інформаційних ресурсів у сучасних умовах часто не дає позитивного результату [5]. Це пов'язано з тим, що теперішні організаційні системи працюють не ізольовано, а в тісній співпраці з іншими Орг.С. Тому забезпечення взаємодії інформаційних систем організацій забезпечує кращу цілісність даних і швидкість передачі інформації.

На рівні файлових систем інтероперабельність забезпечується засобами операційних систем. При цьому досягається певний рівень розподілу доступу, однак така взаємодія обмежується наперед заданими можливостями ОС.

Взаємодія баз даних і сховищ створюється, як правило, в межах конкретних ІКС. При цьому розробляються спеціальні механізми передачі даних між системами. Це пов'язано з тим, що характер інформації в таких типах ресурсів, як правило, передбачає її закритість. Крім того, самі дані та інформація, зазвичай, не мають великої цінності без наявності відповідних методів і

засобів обробки. Тому створення глобальних методів взаємодії баз даних та інформаційних сховищ у загальному вигляді не є доцільним. Розробляються спеціальні алгоритми взаємодії окремих визначених БД і сховищ із забезпеченням необхідного рівня безпеки та максимальної швидкості взаємодії.

Інформаційні колектори розробляються для збереження знань. Знання дуже часто необхідно робити відкритими і загальнодоступними. Тому актуальною є проблема забезпечення інтероперабельності різних колекторів.

Для забезпечення поширення наукових знань у світі розробляються засоби взаємодії наукових архівів. Ініціатива "відкритих архівів" OAI (Open Archives Initiative) [20], започаткована розробниками системи EPrints, створює і поширює стандарти взаємодії та інтероперабельності бібліотечних архівів із метою більш ефективного використання наукових знань. Ця ініціатива включає в себе протокол збору метаданих, оформлених за визначеними стандартами. Ті системи, які підтримують OAI, можуть бути об'єднані в загальну мережу архівів, що взаємодіють через задані інтерфейси.

Для того щоб користувач міг швидко і з однієї точки доступу знайти потрібну йому інформацію, було розроблено ініціативу OCLC (Online Computer Library Center) [22]. Її ідея полягає в тому, що кожному документу ставиться у відповідність набір метаданих, за якими проводиться класифікація і пошук. Таким чином, користувач за своїм запитом може отримати посилання на документи в різних ресурсах різних організацій у всьому світі.

У 2005 р. ідеї OCLC було реалізовано в системі WorldCat, яка забезпечує автоматичну індексацію понад 10000 бібліотечних архівів і містить у собі посилання на більш ніж 1,2 млрд документів [21]. Документи у всіх розглянутих системах можна індексувати у WorldCat. Система CONTENTdm забезпечує автоматичну індексацію документів у каталозі WorldCat, тобто без необхідності будь-яких дій з боку користувача.

Google, Yahoo) не забезпечують користувачу можливості швидко знайти потрібні документи, особливо в разі, коли він не знає точно, що шукає. Однак такі пошукові системи можуть більш-менш ефективно працювати з ресурсами, не задіяними у глобальних проектах типу WorldCat. Саме тому необхідним є поєднання можливостей всіх пошукових систем. Модель глобального інформаційного ресурсу зображено на рис. 2.

Отже, метапошукові і публічні пошукові системи індексують і таким чином упорядковують систему інформаційних ресурсів. Самі ресурси можуть взаємодіяти через стандарти типу OAI, але така взаємодія обмежена порівняно із системами класу WorldCat.

З іншого боку, упорядкована взаємодія колекторів дозволяє створювати системи масштабу країни, такі, наприклад, як “Проект цифрових музеїв Китаю” [23].

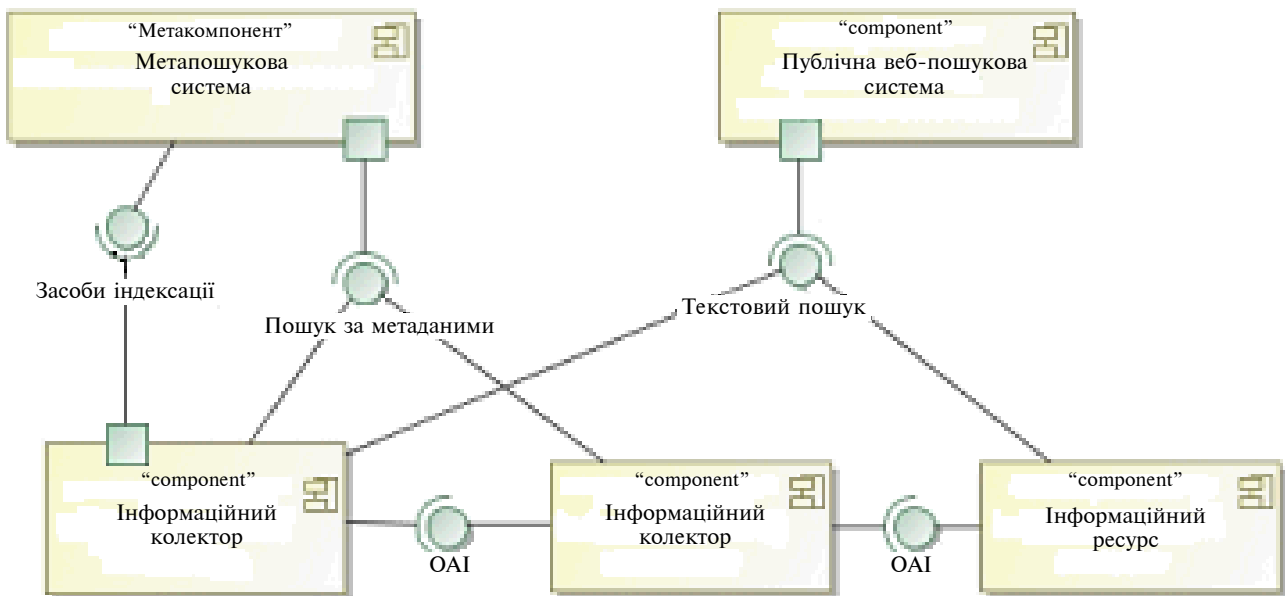


Рис. 2. Модель глобального інформаційного ресурсу

Слід зазначити, що практично всі з наведених функціональних можливостей можуть бути забезпечені і для систем, які не мають базової підтримки таких можливостей. Але ця підтримка за допомогою сторонніх щодо до ресурсу засобів потребує додаткових розробок і витрат.

При врахуванні постійної глобалізації та інтернаціоналізації знань виникає потреба у створенні системи або мережі інформаційних ресурсів. Мережа Інтернет ефективно використовується для поширення наукових знань. При цьому помітно, що сучасні засоби пошуку (Go-

### Продуктування інформаційного ресурсу

Процес продуктування інформаційного ресурсу потребує реалізації кількох стадій виконання робіт [24]:

- проведення бізнес-моделювання Орг.С;
- розробки моделі ІКС;
- формалізації вимог до інформаційного ресурсу;
- проектування ресурсу як системи компонентів (менших ресурсів);

- проектування кожного окремого компонента;
- оцінки і оптимізації проекту;
- реалізації компонентів ресурсу;
- інтеграції компонентів;
- впровадження системи.

Перші дві стадії належать до проектування всієї ІКС. Природно, що ресурс має виконувати певні задачі в межах ІКС, і тому створення його без врахування вимог до ІКС недоцільне. Сам по собі ресурс забезпечує лише обмежені засоби обробки інформації, а отже, так чи інакше потребує інтеграції в ІКС. Тому у випадках, коли ресурси розробляються окремо, їх рідко можна буде ефективно використовувати.

Формалізація вимог до інформаційного ресурсу передбачає опис на певній формальній мові вимог до ресурсу. Такі вимоги можна поділити на функціональні, технічні та економічні. Вони включають у себе опис задач, характеристики продуктивності, надійності, масштабованості, розширюваності, вартість реалізації та підтримки [13].

Проектування ресурсу як системи компонентів передбачає визначення набору і типів ресурсів, які будуть виконувати окремі функції в межах ресурсу ІКС. Необхідно описати їх взаємодію як з іншими компонентами ІКС, так і з зовнішніми інформаційними ресурсами. Для такого моделювання доцільно використати графічну мову моделювання UML та відповідні програмні засоби [3].

Побудова моделей як ресурсу в цілому, так і його компонентів дає змогу розрахувати значення показників ресурсу до початку його реалізації. У випадку, коли характеристики спроектованого ресурсу не відповідають вимогам, поставленим на етапі їх формалізації, необхідно провести оптимізацію проекту, яка передбачає певні зміни на різних рівнях – від рівня найменших компонент до рівня моделі ІКС в цілому. Використання графічних мов моделювання дозволяє за допомогою певних надбудов до систем моделювання відразу після внесення змін у моделі оцінювати нові значення характеристик [24].

Інтеграція компонентів у випадку добре проведеного проектування не викликає проблем і зводиться до налаштування, тестування та часткових виправлень у розроблених компонентах. У процесі впровадження можуть бути помічені недоліки і помилки, допущені на попередніх етапах. Тому процес створення ресурсу є ітеративним.

Процес продукування інформаційного ресурсу як компонента потребує реалізації кількох етапів [24]:

- формалізації задачі продукування інформаційного ресурсу.
- визначення параметрів ресурсу;
- інтеграції інформаційного ресурсу;
- захисту інформаційного ресурсу;
- визначення засобів продукування інформаційного ресурсу;
- системної інтеграції технології продукування інформаційного ресурсу.

На першому етапі розв'язується задача формалізації створення того чи іншого ресурсу. Створення ж інформаційного ресурсу заради самого ресурсу не має сенсу. Інформаційний ресурс повинен сприяти розв'язанню певних задач (існуючих або потенційних). Знаючи задачі, що ставляться перед ресурсом, можна визначити ряд вимог до ресурсу. На даному етапі проводиться формальний опис таких вимог. Він є описом на певній мові вимог до даного ресурсу – від мети його створення до засобів реалізації. Для формалізації може бути використана структурована природна мова, умови, характеристики, моделі, правила тощо.

На другому етапі визначаються конкретні параметри необхідного ресурсу. Серед них можуть бути спосіб і порядок доступу, характеристики інформації (даних, знань), що зберігаються, об'єм даних, структура ресурсу тощо.

Дуже важливою на сьогодні є задача забезпечення захисту інформаційного ресурсу, оскільки в інформаційному суспільстві знання, інформація і дані стають найбільшою цінністю. Все це веде до необхідності врахування питань безпеки при розробці ресурсу на абонентському каналному і мережних рівнях. Для цього використовуються організаційні і технічні методи та засоби захисту. На абонентському рівні застосовуються системи розмежування і розподілу прав доступу до засобів доступу інформаційно-комунікаційної мережі. На каналному рівні використовуються методи сегментації і логічної структуризації інформаційних потоків та відповідні технічні засоби: мережні маршрутизатори та мережні екрани для локальних мереж. На мережному рівні застосовуються міжмережні екрани і криптографічні методи та засоби захисту від несанкціонованого доступу. Для забезпечення захисту інформації, адекватного цінності інформаційного ресурсу, пропонується застосовувати системи керування безпекою організаційної структури, яка об'єднує методи і засоби захисту в єдину систему (Security Manager) [25].

За визначеними на попередніх етапах відомостями вибираються засоби створення інформаційного ресурсу. Створення комп'ютеризованої інформаційно-комунікаційної системи не завжди виправдане. Якщо приймається рішення про доцільність створення саме комп'ютеризованого ресурсу, то для його створення (або для створення окремих його частин) можливе використання існуючих технологій автоматизації розробки баз даних, інформаційних сховищ та колекторів.

Продуктування інформаційного ресурсу являє собою циклічний процес. Після реалізації певного етапу розробки ресурсу виникають нові задачі, які зумовлюють новий цикл вдосконалення. Крім зазначеного порядку роботи, можливий (і природний) зворотний шлях – повернення на попередні етапи продуктування інформаційного ресурсу.

Співвідношення стадій проектування ресурсу, етапів проектування окремого ресурсу та стадій розробки [24, 26] наведено в табл. 1.

- Fedora, яка була першим проектом по створенню універсальної технології продукування інформаційного колектора [27];

- DSpace, яка була розроблена Масачусетським технологічним інститутом (MIT) у співробітництві із Hewlett-Packard для забезпечення інформаційних потреб інституту [28];

- CDS Invenio – розроблена Європейською організацією ядерних досліджень для зберігання ресурсів у галузі ядерних досліджень [29];

- Greenstone – проект новозеландського Проекту цифрових бібліотек і університету Вайкато, призначений для створення простого і компактного ресурсу [30];

- EPrints – розроблений університетом Саутгемптон на основі дослідження вимог освітніх організацій до інформаційного колектора [31];

- Archimede – розроблений бібліотекою університету Лавалю на базі DSpace [32].

Також були розглянуті дві платні системи:

Таблиця 1. Співвідношення стадій продуктування ресурсу, етапів проектування і стадій розробки

Стадії продуктування ресурсу	Етапи проектування ресурсів як компонентів ІКС	Стадії розробки
Проведення бізнес-моделювання Орг.С	–	Аналіз, проектування
Розробка моделі ІКС	–	Аналіз, проектування
Формалізація вимог до інформаційного ресурсу	–	Аналіз, проектування
Проектування ресурсу як системи компонентів (менших ресурсів)	Формалізація Визначення параметрів Інтеграція	Проектування
Проектування кожного окремого компонента	Захист Визначення засобів продукування	Аналіз, проектування
Оцінка і оптимізація проекту	–	Проектування
Реалізація компонентів ресурсу	Системна інтеграція	Проектування, реалізація
Інтеграція компонентів	Системна інтеграція	Реалізація
Впровадження системи	–	Розгортання

### Засоби продукування інформаційних колекторів

На сьогодні на ринку систем продукування інформаційних ресурсів є як безкоштовні, так і платні рішення. Для аналізу було вибрано такі популярні безкоштовні системи:

- Vepress – сервер інформаційних ресурсів, розроблений Berkeley Electronic Press для створення і зберігання інформаційних ресурсів користувачів [33];

- CONTENTdm – розроблений DiMeMa Inc. як інструмент для створення та керування інформаційними архівами з широкими функціональними можливостями [34].

Порівняльний аналіз основних показників існуючих рішень наведено в табл. 2 [19].

Із даної таблиці видно, що в платних системах пропонується краща функціональність, ніж у безплатних. Тому такі системи найбільш прийнятні для компаній, які хочуть швидко і без ризику отримати інформаційний колектор. Державні установи, наукові та освітні організації, які є основними виробниками сучасних знань, вибирають безплатні рішення, які теоретично мають аналогічні можливості, але потребують значних зусиль для налаштування і розгортання.

Практично всі системи зараз підтримують

сучасний стандарт метаданих Dublin Core, текстовий пошук у певному вигляді та засоби веб-доступу. Важливою характеристикою для впровадження систем в Україні є їх багатомовність. При цьому лише одна з розглянутих систем (Greenstone) підтримує російську мову в базовій версії.

Сучасні системи зберігають документи довільного формату. Такі системи, як Fedora мають засоби поєднання мультимедійного контенту із засобами веб-доступу, що дає можливість забезпечити реалізацію гнучкого та зручного інтерфейсу користувача.

Таблиця 2. Порівняльний аналіз основних показників систем продукування інформаційних колекторів

№ п/п	Назва системи	Призначення	Характеристики		
			функціональні	технічні	економічні
1	Fedora	Розробка технологій створення, поширення та повторного використання інформаційних ресурсів, які формують інтелектуальну спадщину людства	Збереження інформації різноманітної природи з довільними зв'язками Збереження метаданих Засоби доступу через "веб"	Метадані Dublin Core Імпорт/експорт XML Масштабованість Розширюваність (API для керування і доступу) OAI-підтримка Версійність контенту Утіліта для "міграції"	Безкоштовна Open source Mozilla open source License
2	DSpace	Репозиторій для збереження, накопичення інформації інституту (institutional repository) (розроблявся для MIT)	Збереження інформації в довільному форматі Взаємодія з навчальними системами Персоналізація веб-інтерфейсу Повнотекстовий пошук (lucene, google) Воркфлоу з прийняття документів Децентралізований процес наповнення	Метадані Dublin Core Імпорт/експорт XML OAI-підтримка Розширюваність через Java API БД: PostgreSQL або транзакційна SQL	Безкоштовна Open source
3	CDS Invenio	Створення електронних бібліотечних каталогів, системи документів, серверу препринтів для великих систем, зокрема ядерної енергетики	Повнотекстовий пошук Персоналізація користувача, кошики документів, повідомлення e-mail	OAI-підтримка Стандарт метаданих MARC 21 БД: MySQL Розширюваність через API Пошук (Google-like) Обмежена підтримка e-mail	Безкоштовна Open source Платна підтримка

Кінець табл. 2

№ п/п	Назва системи	Призначення	Характеристики		
			функціональні	технічні	економічні
4	Eprints	Організація репозитарію препринтів результатів наукових досліджень	Збереження інформації в довільному форматі Повнотекстовий пошук Довільна схема метаданих Веб-інтерфейс Воркфлоу з прийняття документів	БД: MySQL Розширюваність через API (Perl) Генерація RSS	Безкоштовна Open source
6	Greenstone	Створення бібліотечних архівів	Багатомовність (з російською)	Демонстрація перед побудовою колекції. Запис на диск	Безкоштовна Open source GNU GPL
7	Achimedede	Створення інститутських репозитаріїв із багатомовною підтримкою	Продовження Dspace Багатомовність (English, French and Spanish)	Пошук (Lucene) OAI-підтримка Метадані Dublin Core	Безкоштовна Open source
8	Verpress	Створення, зберігання, забезпечення доступу до інформації користувачів у сховищі розробника (Berkeley Electronic Press)	Інструментарій користувача Налаштування через API Повнотекстовий пошук	OAI-підтримка Експорт даних у XML	Повна: 8–50 тис. дол./рік за контент; 4–5 тис. дол. за серію
9	CONTENTdm	Засоби для створення і керування інформаційними сховищами із забезпеченням максимально можливої функціональності	Всі типи контенту Пошук за колекціями Автоматична індексація у WorldCat Поставляється як програмний продукт на комп'ютери користувача	OAI-підтримка Метадані Dublin Core Експорт/імпорт XML Сумісність Z39.50	Повна: 7–40 тис. дол./рік Підтримка: 1–6 тис. дол./рік Інсталяція: 2500 дол.

### Застосування технології

Застосування описаної технології продемонструємо на прикладі побудови ресурсу для Державної податкової адміністрації України (ДПА) [19]. Інформаційно-комунікаційна система ДПА повинна забезпечити облік платників податків, їх доходів та виплат, забезпечити процеси обробки податкових документів, прийом та облік коштів, інформаційні сервіси, такі, як база документів нормативно-правового характеру, надати інструменти аналітичної обробки інформації. Доступ до системи має здійснюватися із локальної мережі ДПА, а також ззовні (через Інтернет, як для співробітників ДПА, так і для громадян). При цьому необхідно створити відповідний рівень безпеки.

Модель ІКС, отриману на етапі проектування, показано на рис. 3.

Портал забезпечує авторизацію та роботу для користувачів і доступ до ресурсів. Платникам податків доступна неавторизована робота через зовнішній портал, основною задачею якого є інформування громадян, відповідь на їх запитання, надання шаблонів документів тощо.

Інформаційний ресурс має надати доступ до оперативних даних, статистичних звітів та документів. Необхідно забезпечити інтеграцію ресурсу з двома групами інформаційних ресурсів: ресурсами інших органів державної влади та публічних ресурсів.

Для опису вимог до системи використовується метрика, яка є необхідною і достатньою множиною характеристик атрибутів і операцій, призначених для проектування та експлуатації ІКС. Для кожного конкретного класу ІКС та її реалізації метрика набуває конкретних значень характеристик.

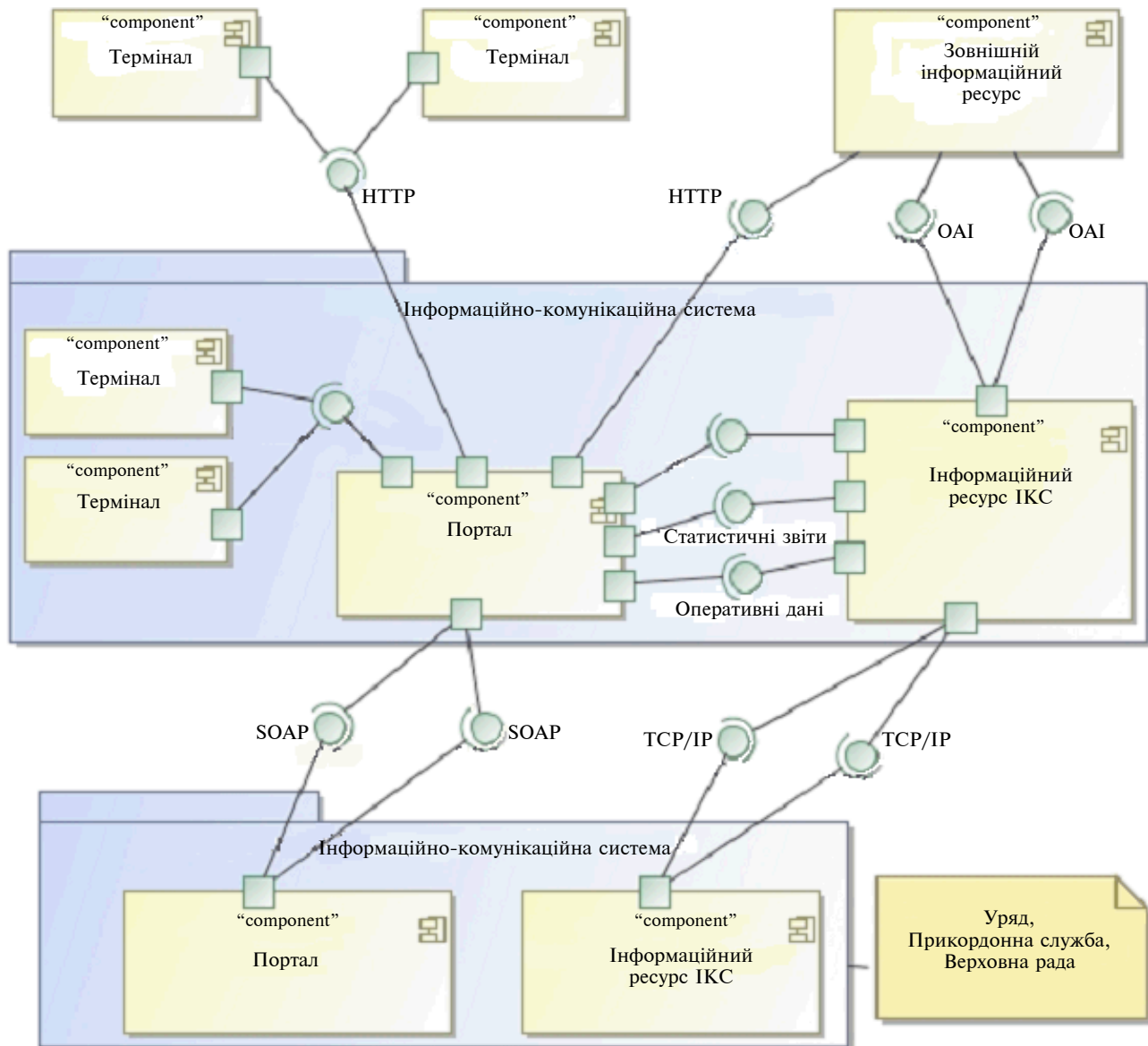


Рис. 3. Модель ІКС Державної податкової адміністрації України. Діаграма компонентів у нотатції UML

Метрика використовується як розробником ресурсу для контролю та покращення якості, так і замовником при оцінці наданого рішення.

Загальноприйнятої метрики для інформаційних ресурсів немає. Це пов'язано з рядом проблем з визначення показників інформаційних ресурсів. Однак розробка такої системи показників є необхідною для створення системи керування та обліку ресурсів [35]. Тому ми пропонуємо включити в метрику найбільш поширені і необхідні для проектування характеристики, які описують атрибути і операції окремих компонентів та/або всього ресурсу. До таких характеристик насамперед слід віднести:

- характеристики продуктивності функціонування (пропускну спроможність, максималь-

но допустимий час відповіді на запит до оперативних даних, аналітичних звітів і документів, час затримки);

- характеристики розширюваності і масштабованості;
- об'єм даних, що зберігаються;
- характеристики безпеки інформаційного ресурсу.

Максимально допустимий час відповіді на запит визначає зручність роботи в системі. Оперативні дані вміщують дані про користувачів системи, реєстр платників податків тощо. Для отримання такої інформації використовуються прості запити до бази даних. Оперативна інформація потрібна практично для кожної дії користувача. Допустимий час її отримання не

має перевищувати одну секунду. Документи і аналітичні звіти необхідні не так часто. Крім того, така інформація має більшу “значимість” для користувача і вимагає певної обробки користувачем, а тому тут допускається більший час відповіді.

Слід зазначити, що час відповіді залежить від кількості одночасно підключених користувачів (терміналів), від чого залежить кількість одночасних запитів. Тому при оцінці часових параметрів треба враховувати навантаження на систему. При цьому варто враховувати також сезонні навантаження на систему, пов’язані з визначеними законодавством періодами подання податкових документів, та загальну кількість платників податків. Так, у Києві кількість платників податків становить 308 тис. фізичних і юридичних осіб. Стійкість і надійність визначається імовірністю відмови, наявністю резервування і часом відновлення системи після збою. Для ДПА допускається тимчасо-

ва недоступність сервісів і не допускається втрата інформації.

Модель інформаційного ресурсу зображено на рис. 4.

Для того щоб забезпечити одночасну обробку запитів від великої кількості операторів, інформаційний ресурс реалізується у вигляді розподіленої системи. Для кожного регіонального відділення податкової адміністрації встановлюється окрема система апаратних і програмних засобів. З погляду на ресурси, до такої регіональної системи належить регіональна база даних, яка зберігає основну інформацію, потрібну для роботи системи в регіонах. Для більшості запитів звернення до центрального серверу не є необхідним.

Центральний сервер зберігає дані та інформацію, потрібні для роботи системи в цілому. До них належать центральна база даних, інформаційний колектор та інформаційне сховище. Інформаційне сховище призначено для збо-

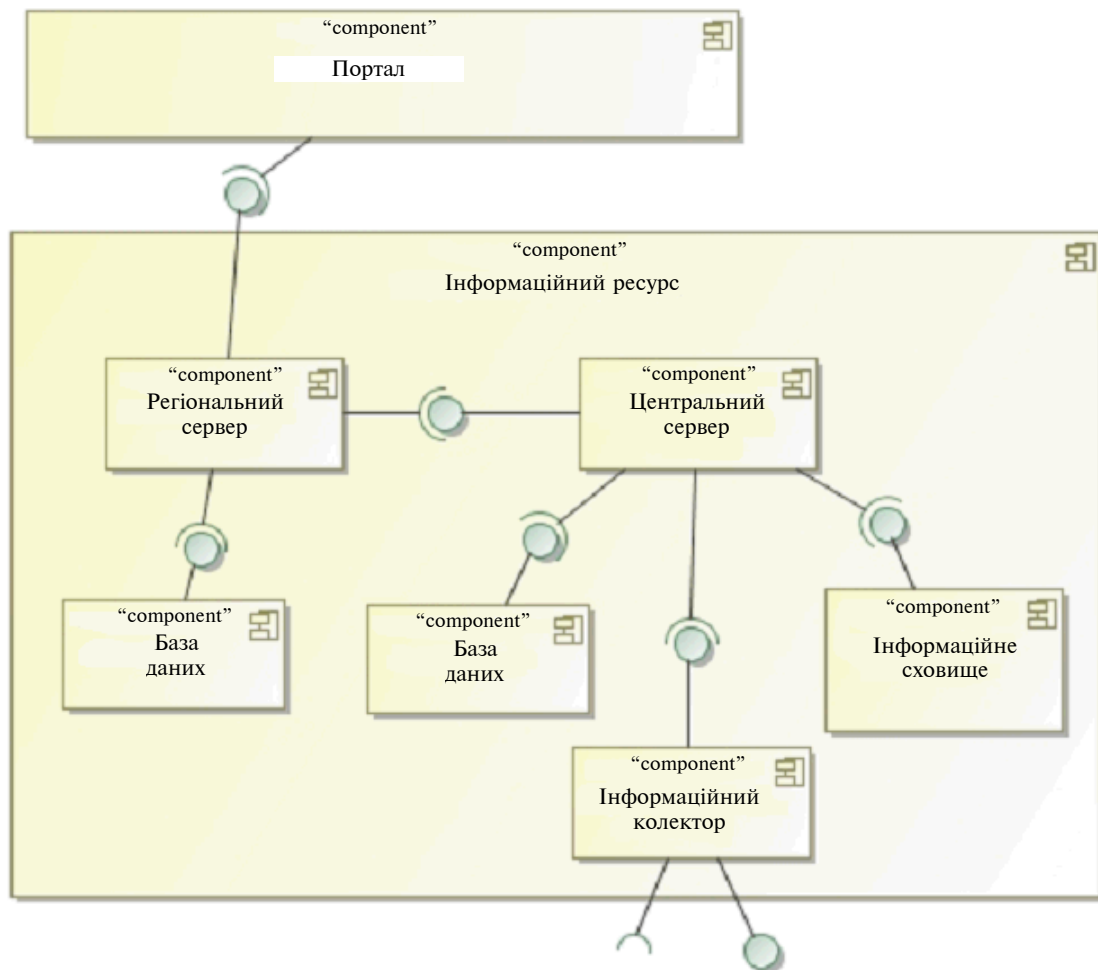


Рис. 4. Модель інформаційного ресурсу ДПА. Діаграма компонентів у нотації UML

ру аналітичної інформації. Наповнення сховища відбувається при заповненні даних у регіональних та центральній БД.

Інформаційний колектор призначено для зберігання нормативних актів, законів, правил, інструкцій та інших документів, на яких базується робота ДПА та які є зручними для користувачів. Колектор забезпечує зберігання інформації ДПА і доступ до інформації в колекторах інших засобів державної влади.

Кожний з компонентів деталізується при подальшому моделюванні. На апаратному рівні як регіональний, так і центральні сервери включають в себе сервер застосувань, сервер баз даних та проксі-сервер. Така система забезпечує необхідний рівень надійності і безпеки.

### Висновки

Застосування прикладного системного аналізу і компонентного підходу до проектування інформаційних ресурсів ІКС дозволяє впорядкувати та істотно спростити процес проектування інформаційного ресурсу, врахувати конкретні вимоги, провести оптимізацію структурного і динамічного подання інформаційного ресурсу, закласти необхідні рішення відповідно до

іменованої специфікації на інформаційний ресурс [24].

Використання для проектування ІКС графічної мови моделювання UML і засобів автоматизації процесу проектування дає змогу автоматично генерувати програмне забезпечення інформаційного ресурсу з його тестуванням на кожній ітерації та фазі розробки.

Таким чином, компонентний підхід дає можливість створити автоматизовану систему проектування ІКС. Складовою такої системи є підсистема автоматизації проектування інформаційного ресурсу, яка дозволяє спроектувати і продукувати ресурс як окремий компонент ІКС.

Специфікація вимог до ресурсу за допомогою визначення іменованих інтерфейсів дозволяє розділити процес розробки ресурсу на незалежні компоненти, які можуть бути спроектовані і розроблені окремо. Визначені вимоги сприяють обґрунтованому вибору засобів створення окремих компонентів ресурсу. Використання сучасних стандартів інтероперабельності забезпечує надійне поєднання, розроблених окремо компонентів інформаційного ресурсу Орг.С на стадіях реалізації та розгортання.

П.П. Маслянюк, П.Н. Лиссов

ИССЛЕДОВАНИЕ СРЕДСТВ И РАЗРАБОТКА ТЕХНОЛОГИИ ПРОДУЦИРОВАНИЯ ИНФОРМАЦИОННЫХ РЕСУРСОВ

Рассмотрено применение прикладного системного анализа и компонентного процесса разработки информационных ресурсов информационно-коммуникационных систем, исследованы определения, классификация и характеристики ресурсов, сделан обзор основных типов ресурсов и их интероперабельности. Предложен способ построения информационного ресурса на базе компонентного подхода. Приведены результаты применения технологии при разработке информационного ресурса Государственной налоговой администрации Украины.

P.P. Maslyanko, P.M. Lissov

TOOLS RESEARCH AND DEVELOPMENT OF THE TECHNOLOGY FOR INFORMATIVE RESOURCES PRODUCTION

The article describes the component approach to production of the informative resources for informative-communicative system. Definitions, classification and main characteristics of the informative resources are presented, and their main types are described. Furthermore, the problems of resources interoperability are studied. The approach to informative resources production, based on the component approach, is suggested. The results of technology application for creation of the informative resources for the State tax administration of Ukraine are studied.

1. *Маслянюк П.П.* Технология информатизации корпоративных структур. Ч. 1 // Корпоративные системы. – 2003. – № 1. – С. 17–19.
2. *Маслянюк П.П.* Технология информатизации корпоративных структур. Ч. 2 // Там же. – № 4. – С. 17–19.

3. *Кватрани Т.* Визуальное моделирование с помощью Rational Rose 2002 и UML / Пер. с англ. – М.: Вильямс, 2003.
4. *Маслянюк П.П., Майстренко О.С.* Моделювання бізнес-процесів організаційної структури // Вісник Східно-

- укр. нац. ун-ту ім. В.І. Даля. – 2007. – № 5(111). – С. 136–141.
5. *Маслянюк П.П., Лісов П.М.* Інформаційні ресурси та засоби їх створення // Там же. – С. 141–145.
  6. *Маслянюк П.П.* Концепція інформатизації корпоративних структур // Наукові вісті НТУУ “КПІ”. – 2003. – № 3. – С. 510–525.
  7. *Ракитина Е.А., Пархоменко В.Л.* Информатика и информационные системы в экономике: Учеб. пособ. Ч. 1. – Тамбов: Изд-во Тамб. гос. техн. ун-та, 2005. – 148 с.
  8. Закон України “Про телекомунікації” (від 18.11.2003 р., № 1280-IV).
  9. [www.answers.com](http://www.answers.com)
  10. *Черненко М., Слепцов С.* Принципы классификации управленческих информационных систем // Корпоративные системы. – 2004. – № 1. – С. 41–46.
  11. Закон України “Про Національну програму інформатизації” (від 16.10.2001 р., № 2684-14).
  12. *Костенко Л.Й., Сорока М.Б.* Бібліотека інформаційного суспільства // Бібліотечн. вісн. – 2002. – № 3. – С. 33–38.
  13. *Маслянюк П.П., Лісов П.М.* Інформаційно-комунікаційні системи та технології обробки інформаційних ресурсів // Вісн. КУЕІТУ “Нові технології”. – 2007. – № 1-2(15-16). – С. 20.
  14. *Маслянюк П.П., Лісов П.М.* Проблеми і технології продукування інформаційних ресурсів // Матер. Всеукр. наук.-практ. конф. “Сучасні тенденції розвитку інформаційних технологій”, 11–13 грудня 2006 р., м. Луганськ. – Луганськ, 2007. – С. 184–189.
  15. [http://ru.wikipedia.org/wiki/Список\\_файловых\\_систем](http://ru.wikipedia.org/wiki/Список_файловых_систем)
  16. *Codd E.* Derivability, Redundancy and Consistency of Relations Stored in Large Data Banks // Com. of the ACM. – 1970.
  17. *Сиротюк О.* Особенности проектирования современных баз данных, available at [www.computerworld.com.ua](http://www.computerworld.com.ua)
  18. *Inmon W.H.* Tech Topic: What is a Data Warehouse? // Prism Solutions. – 1995. – 1. – Р. 9–21.
  19. *Маслянюк П.П., Лісов П.М.* Системне проектування інформаційних ресурсів // Матер. X Міжнар. наук.-техн. конф. “Системний аналіз та інформаційні технології”, 20–24 травня 2008 р., м. Київ. – К., 2008. – С. 105.
  20. *Lagoze C., Sompel H.* The Open Archives Initiative: Building a low-barrier interoperability framework // JCDL '01 June 17–23, 2001. – Roanoke, VA, 2001.
  21. <http://www.worldcat.org>
  22. *Online Computer Library Center*, available at <http://www.oclc.org>
  23. *Tansley R.* Building a Distributed, Standards-based Repository Federation, The China Digital Museum Project // D-Lib Magazine. – 2006. – 12, N 7/8.
  24. *Маслянюк П.П., Лісов П.М.* Дослідження та розробка підсистеми автоматизації проектування інформаційних ресурсів організаційних систем // Матер. II Всеукр. наук.-практ. конф. “Сучасні тенденції розвитку інформаційних технологій”, 8–10 квітня 2008 р., м. Луганськ. – Луганськ, 2008. – С. 57–59.
  25. *Маслянюк П.П.* Концепція управління безпекою інформації в корпоративних структурах // Ювілейна наук.-техн. конф. “Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні”, Київ, 9–11 червня 1998. – К., 1998.
  26. *Маслянюк П.П.* Системне проектування процесів інформатизації // Наукові вісті НТУУ “КПІ”. – 2008. – № 1. – С. 201–208.
  27. <http://www.fedora-commons.org>
  28. <http://www.dspace.org>
  29. <http://cdsware.cern.ch>
  30. <http://www.greenstone.org/cgi-bin/library>
  31. <http://software.eprints.org>
  32. <http://www1.bibl.ulaval.ca/archimede/index.en.html>
  33. <http://www.bepress.com/repositories.html>
  34. <http://contentdm.com>
  35. *Антопольский А.Б.* Проблемы управления публичными информационными ресурсами России // Матер. Междунар. конф. “Программа ЮНЕСКО «Информация для всех; всеобщий доступ к информации»”. Санкт-Петербург, 23–25 июня 2004 г., available at [http://confifap.cpic.ru/upload/spb2004/reports/doklad\\_119.doc](http://confifap.cpic.ru/upload/spb2004/reports/doklad_119.doc)

Рекомендована Радою навчально-наукового комплексу “Інститут прикладного системного аналізу” НТУУ “КПІ”

Надійшла до редакції  
20 червня 2008 року