

Дослідження методів та розробка технології продукування інформаційних ресурсів

Маслянюк П.П., Лісов П.М.,

В статті розглядається компонентний підхід до побудови інформаційних ресурсів інформаційно-комунікаційних систем. Розглядається визначення, класифікація та характеристики ресурсів, проведено огляд основних типів ресурсів та їх інтероперабельності. Запропоновано підхід до продукування інформаційного ресурсу на базі компонентного підходу. Представлені результати застосування технології до розробки інформаційного ресурсу Державної податкової адміністрації України.

ВСТУП

Сьогодні одним із ефективних засобів підвищення конкурентоздатності організаційної системи (Орг.С.) є її інформатизація. Під визначенням «Інформатизація організаційних систем» ми 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 та автоматизації бізнес-процесів, підтримки прийняття рішень та підвищення ефективності управління організаційною системою із застосуванням інформаційно-комунікаційних технологій [1, 2, 3].

Інформатизація передбачає створення інформаційно-комунікаційної системи (ІКС) [4]. Однією із компонентів інформаційно-комунікаційної системи є інформаційний ресурс. Він забезпечує зберігання та обробку даних, інформації та знань, які сьогодні є однією 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, фірми прямо залежить від кількості та якості інформації, якою вона володіє, а також від наявності засобів оперування такою інформацією. Тому проблема створення інформаційних ресурсів є важливою для конкурентоспроможності, а отже у сучасних умовах існування організаційної системи.

ПОСТАНОВКА ЗАВДАННЯ

Метою роботи є дослідження методів та розробка технології продукування інформаційних ресурсів. Необхідно запропонувати підхід, що дозволить обґрунтовано обрати засоби та технології для створення інформаційних ресурсів.

Для цього необхідно проаналізувати дані, інформація та знання, які можуть зберігатись в інформаційних ресурсах.

На основі компонентного підходу та прикладного системного аналізу необхідно розробити підхід до проектування та 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 роботи спрямовані на застосування при розробці проектів інформатизації організаційних систем, зокрема для системного проектування інформаційних ресурсів.

ІНФОРМАЦІЙНИЙ РЕСУРС

Однозначного та завжди коректного визначення понять дані, інформація, знання, нажаль, немає. Сам термін “інформація” є порівняно молодим: до II половини XX сторіччя в російській та українській мовах поняття “інформація” не вживалось [6]. У перекладі з латини *informatio* – відомості, роз'яснення. Можна навести багато визначень інформації, від філософського “Відображене різноманіття, яке виникає в результаті взаємодії між об'єктами” [7] до визначення Лебедєва – “послідовність нулів та одиниць”. Однак перше визначення не є конструктивним а друге не зовсім відповідає сучасному розумінню інформації.

Згідно законодавства України інформація – відомості, подані у вигляді сигналів, знаків, звуків, рухомих або нерухомих зображень чи в інший спосіб [8]. Законодавство України визначає дані як інформація 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 та досвід експерта, що дає вагомий результат, який може бути використано для прийняття рішень. Знання може бути вичерпним та/або вузьким, індивідуальним та/або колективним. Нажаль, українське законодавство взагалі не визначає поняття “знання”.

Згідно інших джерел, “Data is information before it has been given any context, structure and meaning” [9] – Дані це інформація до того, як їм надається контекст, структура та значення.

Найбільш вдалим, на мою думку, є визначення, за яким дані – це результат простого збору визначених фактів; інформацією вони стають лише при зв’язуванні у щось корисне, комбінацію хто, що, де і як. У свою чергу знання – це розуміння, як і чому щось відбувається [10]. Саме такі значення понять “дані”, “інформація” та “знання” будуть використовуватись далі.

Дані, інформація та знання – сутності, які можуть відображати об’єкти, процеси та явища навколишнього середовища, вони є абстрактними об’єктам [6]. Засобом їх матеріалізації є інформаційний ресурс. Згідно законодавства України інформаційний ресурс – сукупність документів у інформаційних системах (бібліотеках, архівах, банках даних тощо) [11].

Документ – це упорядкована сукупність даних, інформації та знань, яка надає можливості доступу, передачі, обробки, тощо. Прикладом документа може бути паперовий документ, фільм, комп’ютерний файл, тощо [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, інформаційний колектор.

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

Для опису вимог до інформаційного ресурсу, що розробляється, ми розглядаємо ряд його характеристик [13]. Класифікація дозволяє розбити ресурс на компоненти, визначити характеристики, обрати оптимальні засоби для їх реалізації.

1. Характеристики продуктивності:

- пропускна спроможність;
- час реакції;
- час затримки.

2. Характеристики надійності.

3. Характеристики масштабованості.

4. Характеристики розширюваності.

5. Характеристики прозорості.

6. Повна вартість володіння.

Повна вартість володіння інформаційного ресурсу визначається не тільки (і не стільки) вартістю використаних при його створенні апаратних і програмних засобів, а і вартістю інформації, яка в нього закладена.

Класифікація інформаційних ресурсів дозволяє обґрунтовано визначити необхідний тип інформаційного ресурсу, обрати інструменти та технології для його продуктування.

Інформаційні ресурси можна класифікувати за такими ознаками [14].

1. Приналежність ресурсу до певної організаційно-технологічної системи (наприклад, бібліотечної мережі, ЗМІ, корпоративної системи).
2. Спосіб виділення об’єктів обліку (твори, документи, видання, бази даних, інтернет-сторінки, сайти, тощо).
3. Призначення ресурсу (масова інформація, освіта, бізнес, особиста переписка, тощо).
4. Зміст ресурсу:
 - тематичний;
 - об’єктний;
 - функціональний.
5. Видовий склад ресурсу (види документів).
6. Джерело інформації:
 - національне або закордонне;

- офіційне або неофіційне;
- тощо.
- 7. Правовий статус ресурсу (публічні документи, об'єкти інтелектуальної власності, спам, таємні документи, тощо).
- 8. Структурний тип ресурсу, що включає:
 - можливість відділення даних від програм та представлення;
 - формати;
 - кодування;
 - інше.
- 9. Відкритість ресурсу (відкритий або з обмеженим доступом).
- 10. Рівень структурованості:
 - структуровані;
 - неструктуровані.
- 11. Спосіб розповсюдження і носій.
- 12. Мова ресурсу.

Комп'ютеризовані інформаційні **ресурси за (критерій???) рівнем** надання сервісів користувачу можна розділити на такі основні типи [13]:

1. Файлові системи.
2. Бази даних.
3. Інформаційні сховища.
4. Інформаційні колектори.
5. Веб-ресурси.

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

Розвиток ФС почався із створенням носіїв даних достатньо великого розміру, та виникнення задач, які вже не дозволяли зберігати на носії дані одного типу у визначеному форматі. Можна виділити такі типи ФС [15]:

1. Дискові. Орієнтовані на зберігання розрізаних даних на одному жорсткому диску персонального комп'ютера користувача. Здебільшого використовуються на ПК для вирішення персональних задач користувача.
2. Кластерні. Орієнтовані на зберігання даних на масиві дисків, що забезпечує вищу швидкість, надійність, ємність. Як правило використовуються в серверах, хоча все частіше застосовуються в ПК (RAID-масиви).
3. Мережеві. Дозволяють пов'язати в єдину ФС носії на різному апаратному забезпеченні, поєднані в межах мережі. Розвиток таких систем дозволяє створити інформаційний ресурс підприємства із забезпеченням максимальної надійності і швидкості роботи при уникненні дублювання.
4. Спеціального призначення. До цієї групи відносяться ФС, оптимізовані для роботи на різноманітних носіях (таких як компакт-диски), системи доступу до ftp, системи файлових пристроїв, тощо.

Ми розглядаємо два основні класи задач, які вирішує ФС: організація розміщення даних на фізичному носії та організація файлового простору користувача (ієрархії файлів та каталогів).

Фізична організація зберігання даних на диску передбачає, що для користувача (програми) створюється інтерфейс за певними правилами, тобто певний образ організації файлів і каталогів. Реальне місце кожного файлу встановлюється ФС за певним алгоритмом. Такий алгоритм повинен забезпечити швидкий доступ до файлів, пошук, запис, операції переміщення/копіювання, тощо. Це завдання перш за все 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 та зручність її обробки, підтримку стандартних рішень. Як правило ієрархію файлів кожен користувач визначає самостійно, але інколи існують рішення, орієнтовані на певну визначену ієрархію. Прикладом може бути організація ФС для ОС Windows. Користувач може створювати ієрархію для своїх даних але для ОС використовуються ієрархія, створена розробником. Аналогічно, якщо у межах організаційної структури використовуються певні загальні програмні рішення, може з'явитись необхідність розробки ієрархії ФС. Така розробка має проводитись на етапі проектування, а в подальшому за допомогою організаційних засобів має контролюватись підтримка встановленої ієрархії.

Останнім часом на файлові системи покладаються нові задачі, такі як контроль доступу до даних, підтримка політик безпеки, архівація, тощо. Це викликано тим, що ФС у сучасних системах вже не є лише основою операційної системи (ОС) а являє собою окремих інформаційний ресурс, доступ до якого може проводитись і “в обхід” ОС. Тому функції, які раніше покладались на ОС, тепер мають бути перекладені на саму ФС.

База даних (БД) являє собою інформаційний ресурс, призначений для зберігання структурованих даних. Класифікація БД за структурою інформації, що зберігається:

1. Каталоги. Найпростіший вид БД у якому використовується підхід, аналогічний до бібліотечних каталогів – вся інформація зберігається у одній таблиці з визначеним набором полів. Такі БД були історично першими і почали розвиватись у 1960-х.
2. Ієрархічні. У таких БД дані організовані у деревовидну структуру, у якій кожен запис є вузлом. Така організація БД забезпечує швидкий пошук даних (за одним критерієм).
3. Мережеві. Такі БД зберігають дані із зв'язками з іншими даними. Зв'язки організуються як вказівники (pointer) на інші записи. Як правило для зберігання записів організуються схожа до ієрархічної схема.
4. Реляційні. Найпопулярніші сучасні БД, які базуються на реляційній моделі, запропонованій Е. Коддом у 1969 р. [16]. У таких БД дані представлені у вигляді таблиць, кожен рядок яких представляє один запис а кожний стовпчик – визначений атрибут. Робота з БД відбувається за допомогою запитів, результатом яких є таблиці.
5. Пост-реляційні. Такі бази використовують деякі принципи реляційних БД але не обмежуються реляційною моделлю. Часто вони використовують принципи ієрархічних та мережевих БД.
6. Об'єктні. Враховуючи розвиток об'єктно-орієнтованого програмування виконуються спроби створити схему БД, яка б дозволила працювати з об'єктами. В такому випадку БД і програма, яка її використовує, мають єдині імена і це дозволяє спростити взаємодію і уникнути проблем із перетвореннями даних.

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

Класифікація баз даних за характером інформації, яка зберігається:

- фактографічні (картотеки),
- документальні (архіви);

Класифікація баз даних за за способом зберігання даних:

- централізовані (зберігаються на одному комп'ютері),
- розподілені (використовуються в локальних і глобальних комп'ютерних мережах);

Інформаційні сховища являють собою ресурси, призначені для аналітичної обробки інформації. За визначенням Білла Інмона, сховище даних – це предметно-орієнтований, хронологічний, стабільний набір даних, призначений для підтримки прийняття управлінських рішень [18]. З іншого боку, інформаційне сховище являє собою ієрархічно організовану сукупність БД, призначених для збереження та обробки архівних даних.

Бази даних дозволяють ефективно працювати із оперативними даними, але для аналітичних запитів час обробки може бути великим, так як доводиться перероблювати великі об'єми інформації. Натомість у інформаційне сховище додаються оброблені певним чином дані, тобто первинна аналітична обробка проводиться постійно і поступово. Таким чином отримується багатовимірна структура, яка не містить оперативних даних але містить сировину для аналітичної роботи. За запитом експерта сховище дозволяє швидко побудувати так званий “зріз”, який містить необхідну інформацію.

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

Першими інформаційними колекторами були бібліотечні системи. Завданням таких систем було зберігання текстів документів у сховищах бібліотек [12]. Документи були класифіковані за стандартними бібліотечними класифікаторами. В сучасних інформаційних колекторах для пошуку документів використовуються метадані [20]. Це дозволяє більш ефективно та швидко знаходити документи.

У сучасних умовах дуже важливим стало завдання інтеграції колекторів. Це пов'язано з характером знань, що зберігаються у колекторах [19]. Більшість колекторів зберігає саме відкриту інформацію. Створення єдиної мережі колекторів [21, 22] дозволить забезпечити можливості пошуку інформації у різних колекторах без необхідності проводити пошук у кожному окремо.

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

Веб-ресурси представляють собою ресурси, які зберігаються на різних компонентах, розподілених у гетерогенній мережі. Прикладом такого ресурсу є мережа Інтернет. Перевагою даного типу ресурсу є надійність та забезпечення зручного доступу. Недоліком є складність організації управління і пошуку.

Як видно з аналізу, кожен з розглянутих типів інформаційних ресурсів призначається для вирішення певних завдань. При побудові ІКС Орг.С. необхідно вирішувати різні типи завдань, тому у більшості випадків використовується комбінований ресурс, що поєднує у собі систему окремих взаємодіючих ресурсів. Компонентний підхід дозволяє описати такі ресурси та взаємодію між ними на рівні інтерфейсів. Типова модель організації ресурсів у складі інформаційного ресурсу ІКС представлена на рис. 1.

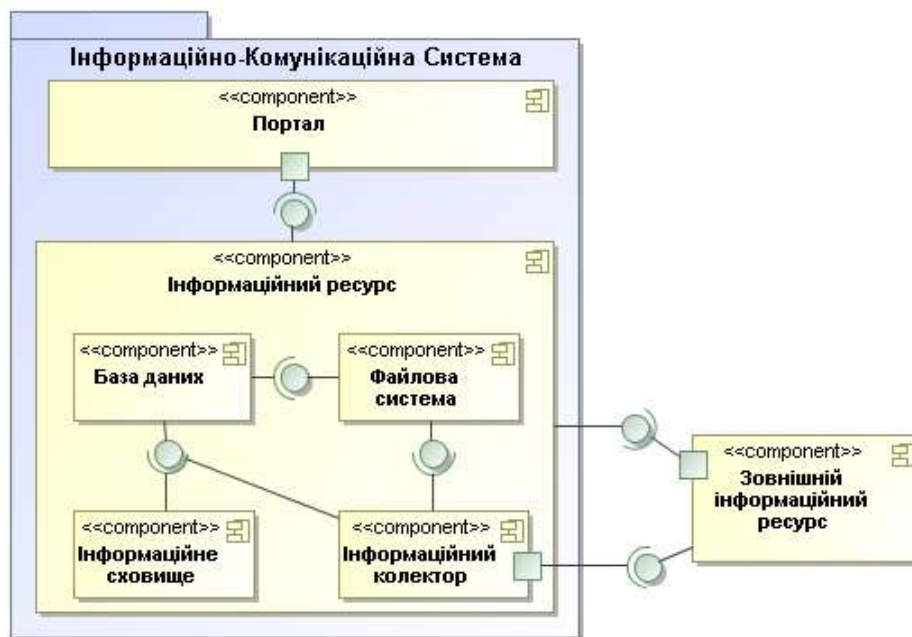


Рисунок 1 – Модель організації інформаційного ресурсу ІКС. Діаграма компонентів у нотації UML

На рисунку показані інтерфейси взаємодії як із зовнішніми для ресурсу компонентами (портал, зовнішній ресурс) так і між внутрішніми компонентами (ресурсами) у межах комплексного ресурсу.

ІНТЕГРАЦІЯ І ІНТЕРОПЕРАБЕЛЬНІСТЬ ІНФОРМАЦІЙНИХ РЕСУРСІВ

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

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

Взаємодія між базами даних та сховищами організується, як правило, в межах конкретних ІКС. При цьому розробляються спеціальні механізми передачі даних між системами. Це пов'язано з тим, що характер інформації у таких типах ресурсів як правило передбачає її закритість. Крім того самі дані та інформація як правило не представляють великої цінності без наявності відповідних методів та засобів обробки. Тому створення глобальних методів взаємодії між базами даних та інформаційними сховищами у загальному вигляді не є доцільним. Розробляються спеціальні алгоритми взаємодії окремих визначених БД та сховищ із забезпеченням необхідного рівня безпеки та максимальної швидкості взаємодії.

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

Для забезпечення розповсюдження наукових знань у світі розробляються засоби взаємодії наукових архівів. Ініціатива “Відкриті архіви” OAI (Open Archives Initiative) [20], започаткована розробниками системи EPrints, розробляє і просуває стандарти взаємодії та інтероперабельності бібліотечних архівів з метою більш ефективного використання наукових знань. Ініціатива включає в себе протокол збору метаданих, оформлених за визначеними стандартами. Ті системи, які підтримують OAI, можуть бути поєднані у загальну мережу архівів, які взаємодіють через задані інтерфейси.

Для того, щоб користувач міг швидко і з одного місця знайти потрібну йому інформацію була розроблена ініціатива OCLC (Online Computer Library Center) [22]. Її ідея полягає у 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 у різних ресурсах різних організацій по всьому світу.

У 2005 році ідеї OCLC були реалізовані в системі WorldCat, яка забезпечує автоматичну індексацію більше 10000 бібліотечних архівів і містить посилання на більш ніж 1.2 млрд. документів [21]. Документи у всіх розглянутих системах можуть бути індексовані у WorldCat. Система CONTENTdm забезпечує автоматичну індексацію документів у каталозі WorldCat, тобто без необхідності яких-небудь дій зі сторони користувача.

Варто зазначити, що практично всі із зазначених функціональних можливостей можуть бути забезпечені і для систем, які не мають базової підтримки таких можливостей. Але така підтримка за допомогою сторонніх по відношенню до ресурсу засобів потребує додаткових розробок та витрат.

Враховуючи постійну глобалізацію та інтернаціоналізацію знань виникає потреба у створенні системи або мережі інформаційних ресурсів. Мережа Інтернет ефективно використовується для розповсюдження наукових знань. При цьому не можна не помітити, що сучасні засоби пошуку (такі як Google, Yahoo) не забезпечують користувачу можливості швидко знайти потрібні документи, особливо у випадку коли користувач не знає точно що він шукає. Однак такі пошукові системи можуть більш-менш ефективно працювати з ресурсами, які не приймають участь в глобальних проектах типу WorldCat. Саме тому необхідним є поєднання можливостей всіх пошукових систем. Модель глобального інформаційного ресурсу представлена на рис. 2.

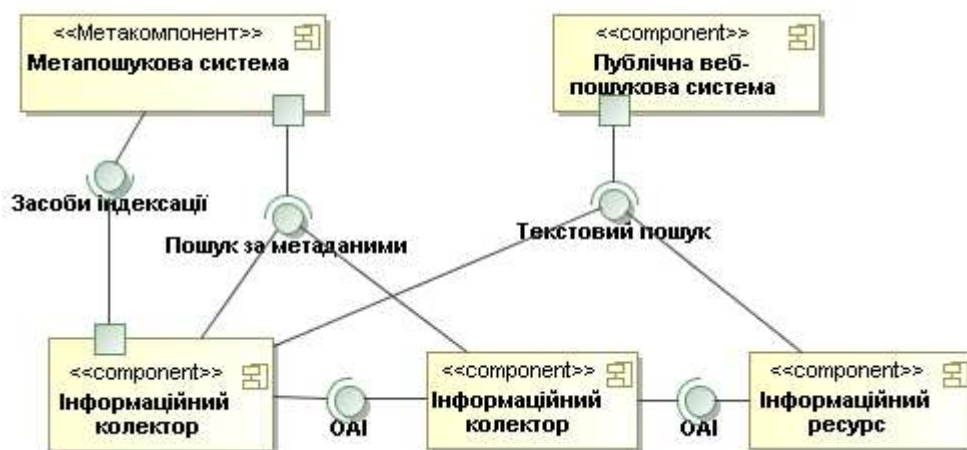


Рисунок 2 – Модель глобального інформаційного ресурсу

Таким чином метапошукові та публічні пошукові системи індексують і таким чином упорядковують систему інформаційних ресурсів. Самі ресурси можуть взаємодіяти через стандарти типу OAI, але така взаємодія є обмеженою у порівнянні з системами класу WorldCat. З іншого боку упорядкована взаємодія колекторів дозволяє створювати системи масштабу країни, такі як Проект цифрових музеїв Китаю [23].

ПРОДУКУВАННЯ ІНФОРМАЦІЙНОГО РЕСУРСУ

Процес продукування інформаційного ресурсу потребує реалізації декількох (робіт?) стадій [24]:

1. Проведення бізнес-моделювання Орг.С.
2. Розробка моделі ІКС.
3. Формалізація вимог до інформаційного ресурсу.
4. Проектування ресурсу як системи компонентів (менших ресурсів).
5. Проектування кожного окремого компоненту.
6. Оцінка та оптимізація проекту.
7. Реалізація компонентів ресурсу.
8. Інтеграція компонентів.
9. Впровадження системи.

Перші дві стадії відносяться до проектування всієї ІКС. Природно, що ресурс повинен виконувати певні задачі в межах ІКС, тому створення його без урахування вимог до ІКС не доречно. Сам по собі ресурс забезпечує лише обмежені засоби обробки інформації, а тому так чи інакше потребує інтеграції у ІКС. Тому у випадках, коли ресурси розробляються окремо, вони рідко можуть бути ефективно використані.

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 опис задач, характеристики продуктивності, надійності, масштабованості, розширюваності, вартість реалізації та підтримки [13].

Проектування ресурсу як системи компонентів передбачає визначення набору та 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 доречно використати графічну мову моделювання UML та відповідні програмні засоби [3].

Побудова моделей як ресурсу в цілому так і його компонентів дозволяє розрахувати значення показників ресурсу до початку його реалізації. У випадку, якщо характеристики спроектованого ресурсу не відповідають вимогам, поставленим та етапі формалізації вимог, 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 надстроек до систем моделювання одразу 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. В процесі впровадження можуть бути знайдені недоліки та помилки, допущені на попередніх етапах. Тому процес створення ресурсу є ітеративним.

Процес продукування інформаційного ресурсу як компоненти потребує реалізації декількох етапів [24].

1. Формалізація задачі продукування інформаційного ресурсу.
2. Визначення параметрів ресурсу.
3. Інтеграція інформаційного ресурсу.
4. Захист інформаційного ресурсу.
5. Визначення засобів продукування інформаційного ресурсу.
6. Системна інтеграція технології продукування інформаційного ресурсу.

На першому етапі вирішується задача формалізації створення того чи іншого ресурсу. Створення інформаційного ресурсу заради самого ресурсу не має сенсу. Інформаційний ресурс повинен дозволяти вирішувати певні задачі (існуючі або потенціальні). Знаючи задачі, що ставляться перед ресурсом, можна визначити ряд вимог до ресурсу. На даному етапі проводиться формальний опис таких вимог. Формальний опис представляє собою опис 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 та порядок доступу, характеристики інформації (даних, знань), що зберігаються, об'єм даних, структура ресурсу, тощо.

Дуже важливою на сьогодні є задача забезпечення захисту інформаційного ресурсу, тому що у інформаційному суспільстві знання, інформація, дані стають найбільшою цінністю. Все це призводить до необхідності урахування питань безпеки при розробці ресурсу на абонентському каналному і мережених рівнях. Для цього використовуються організаційні і технічні методи та засоби захисту. На абонентському рівні застосовуються системи розмежування та розподілу прав доступу до засобів доступу інформаційно-комунікаційної мережі. На каналному рівні

застосовуються методи сегментації та логічної структуризації інформаційних потоків та відповідні технічні засоби, як то мережеві маршрутизатори та мережеві екрани для локальних мереж. На мережевому рівні застосовуються міжмережеві екрани та криптографічні методи та засоби захисту від несанкціонованого доступу. Для забезпечення адекватної цінності інформаційного ресурсу захисту інформації пропонується застосування системи управління безпекою організаційної структури, яка об'єднує методи і засоби захисту в єдину систему (Security Manager) [25].

За визначеними на попередніх етапах відомостями обираються засоби створення інформаційного ресурсу. Створення комп'ютеризованої інформаційно-комунікаційної системи не завжди є виправданим. 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 інформаційних колекторів.

Продуктування інформаційного ресурсу являє собою циклічний процес. Після реалізації певного етапу розробки ресурсу виникають нові задачі, які призводять до необхідності нового циклу вдосконалення. Крім зазначеного порядку роботи можливий (і природний) зворотній шлях – повернення на попередні етапи продуктування інформаційного ресурсу.

Співвідношення стадій проектування ресурсу, етапів проектування окремого ресурсу та стадій розробки [24, 26] наведено у табл. 0.1.

Таблиця 0.1. Співвідношення стадій продуктування ресурсу, етапів продуктування та стадій розробки

Стадії продуктування ресурсу	Етапи проектування ресурсів як компонентів	Стадії розробки
Проведення бізнес-модельювання 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 компоненту	Захист Визначення засобів продуктування	Аналіз, проектування
Оцінка та 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 продуктування інформаційних ресурсів представлені як безкоштовні так і платні рішення. Для аналізу були обрані такі популярні безплатні системи:

- Fedora, яка була першим проектом по створенню універсальної технології продуктування інформаційного колектору [27];
- DSpace, 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, який був розроблений університетом Саутгемптону на основі дослідження вимог освітніх організацій до інформаційного колектору [31];
- Archimede, розроблений бібліотекою університету Лавалля на базі DSpace [32].

Також були розглянуті 2 платні системи:

- Verpress, сервер інформаційних ресурсів, розроблений Berkeley Electronic Press для створення та зберігання інформаційних ресурсів користувачів [33];
- CONTENTdm, розроблений DiMeMa, Inc. як інструмент для створення та управління інформаційними архівами з широкими функціональними можливостями [34].

Порівняльний аналіз основних показників існуючих рішень наведено у табл. 1 [19].

Таблиця 1. Порівняльний аналіз основних показників систем продукування інформаційних колекторів

№ п/п	Назва системи	Призначення	Функціональні характеристики	Технічні характеристики	Економічні характеристики
1	Fedora	Розробка технологій розробки, створення, розповсюдження 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 Маштабованість Розширюваність (API для управління і доступу) ОАІ-підтримка Версійність контенту Утіліта для "міграції"	Безкоштовна Open source Mozilla open source License
2	DSpace	Репозиторій для збереження, накопичення, .. Інформації інституту (intitutional repository) (розроблявся для MIT)	Збереження інформації у довільному форматі Взаємодія з навчальними системами Персоналізація веб-інтерфейсу Повнотекстний пошук (lucene, google) Воркфлоу по прийняттю документів Децентр. процес наповн.	Метадані Dublin Core Імпорт/експорт XML ОАІ-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	Повнотекстний пошук Персоналізація користувача, кошики документів, e-mail повідомлення	ОАІ-підтримка Стандарт метаданих MARC 21 БД: MySQL Розширюваність через API Пошук (Google-like) Обмеж. підтримка e-mail	Безкоштовна Open source Платна підтримка
4	Eprints	Організація репозитарію препринтів результатів наукових досліджень	Збереження інформації у довільному форматі Повнотекстний пошук Довільна схема метаданих Веб-інтерфейс Воркфлоу по прийняттю 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	Achimede	Створення інститутських репозитаріїв з багатомовною підтримкою	Продовження Dspace Багатомовність (English, French and Spanish)	Пошук (Lucene) ОАІ-підтримка Метадані Dublin Core	Безкоштовна Open source
8	berpress	Створення, зберігання, забезпечення доступу, до інформації користувачів у сховищі розробника (Berkeley Electronic Press)	Інструментарій користувача Налаштування через API Повнотекстний пошук	ОАІ-підтримка Експорт даних у XML	Повна: 8-50 тис. дол./ рік за контент: 4-5 тис. дол. За серію
9	CONTENT dm	Засоби для створення і управління інформаційними сховищами із забезпеченням максимально можливої функціональності	Всі типи контенту Пошук по колекціям Автоматична індексація у WorldCat Поставляється як програмний продукт на комп'ютери користувача	ОАІ-підтримка Метадані Dublin Core Експорт/імпорт XML Сумісність Z39.50	Повна: 7-40 тис. дол./ рік Підтримка: 1-6 тис. дол. на рік Інсталяція: 2500 дол.

Із таблиці видно, що платні системи пропонують кращу функціональність, ніж безкоштовні. Тому такі системи найбільш прийнятні для компаній, які хочуть швидко та без ризику отримати інформаційний колектор. Державні установи, наукові та освітні організації, які є основними виробниками сучасних знань, обирають безкоштовні рішення, які теоретично мають аналогічні можливості, але потребують значних зусиль для налаштування і розгортання.

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

Сучасні системи підтримують зберігання документів довільного формату. Такі системи, як Fedora мають засоби поєднання мультимедійного контенту із засобами веб-доступу, що дозволяє забезпечити реалізацію гнучкого та зручного інтерфейсу користувача.

ЗАСТОСУВАННЯ ТЕХНОЛОГІЙ

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, забезпечення процесів обробки податкових документів, забезпечення прийому та обліку коштів, забезпечити інформаційні сервіси, такі як базу документів нормативно-правового характеру, надати інструменти аналітичної обробки інформації. Доступ до системи має забезпечуватись із локальної мережі ДПА а також ззовні (через Інтернет, як для співробітників ДПА так і для громадян). При цьому необхідно забезпечити відповідний рівень безпеки.

Модель ІКС, отримана на етапі проектування, показана на рис. 3.

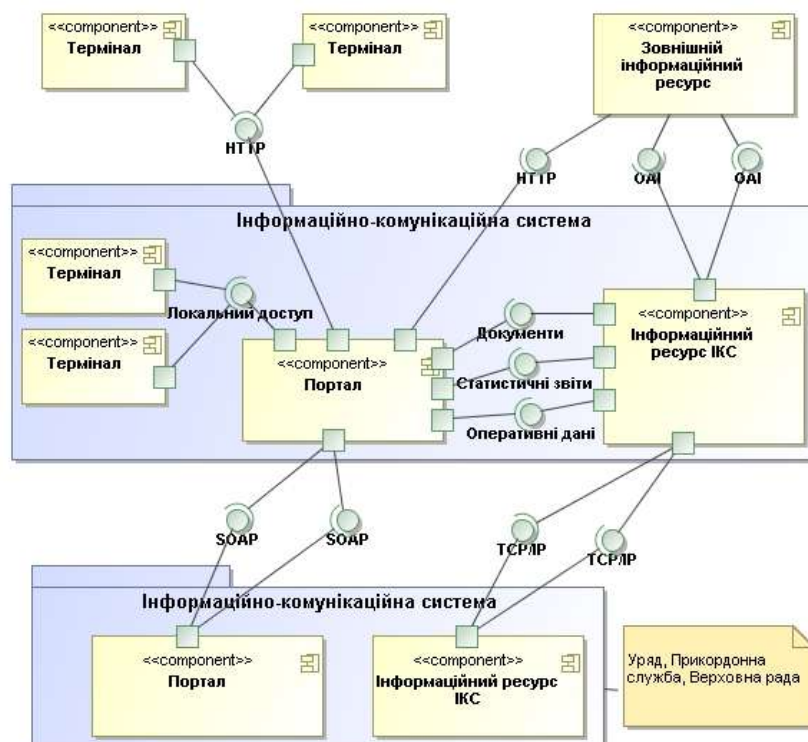


Рисунок 3 – Модель ІКС Державної податкової адміністрації України. Діаграма компонентів в нотації UML

Портал забезпечує авторизацію та забезпечення роботи користувачів, доступ до ресурсів. Для платників податків забезпечується неавторизована робота через зовнішній портал, основною задачею якого є інформування громадян, відповідь на їх запитання, представлення шаблонів документів, тощо.

Інформаційний ресурс має забезпечити 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. Метрика – це необхідна і достатня множина характеристик атрибутів і операцій, призначених для проектування та експлуатації ІКС. Для кожного конкретного класу ІКС та її реалізації метрика набуває конкретних значень характеристик.

Метрика використовується як розробниками ресурсу для контролю та покращення якості так і замовником при оцінці представленого рішення.

Загальноприйнятої метрики для інформаційних ресурсів немає. Це пов'язано з рядом проблем по визначенню показників інформаційних ресурсів. Однак розробка такої системи показників є необхідною для створення системи управління та обліку ресурсів [35].

Тому ми пропонуємо включити в метрику найбільш розповсюджені та необхідні для проектування характеристики, які описують атрибути та операції окремих компонентів та/або всього ресурсу.

До таких характеристик насамперед слід віднести:

- характеристики продуктивності функціонування (пропускну спроможність, максимально допустимий час відповіді на запит до оперативних даних, аналітичних звітів та документів);
- характеристики розширюваності та масштабованості;
- об'єм даних, що зберігаються;
- характеристики безпеки інформаційного ресурсу.

Максимально допустимий час відповіді на запит визначає зручність роботи у системі. Оперативні дані включають дані про користувачів системи, реєстр платників податків, тощо. Для отримання такої інформації використовуються прості запити до бази даних. Оперативна інформація необхідна практична для кожної дії користувача. Допустимий час її отримання не має перевищувати 1с. Документи та аналітичні звіти необхідні не так часто. Крім того така інформація представляє більшу “значимість” для користувача і вимагає певної обробки користувачем, тому тут допустимий більший час відповіді.

Необхідно зазначити, що час відповіді залежить від кількості одночасно підключених користувачів (терміналів), від якої залежить кількість одночасних запитів. Тому при оцінці часових параметрів необхідно враховувати навантаження на систему. При цьому варто врахувати сезонні навантаження на систему, пов'язані з визначеними законодавством періодами подання податкових документів та загальну кількість платників податків. Так у Києві кількість платників податків складає 308 тисяч фізичних і юридичних осіб. Стійкість та надійність визначається імовірністю відмови, наявністю резервування та часом відновлення системи після збою. Для ДПА допускається тимчасова недоступність сервісів і не допускається втрата інформації.

Модель інформаційного ресурсу представлено на рис.4.

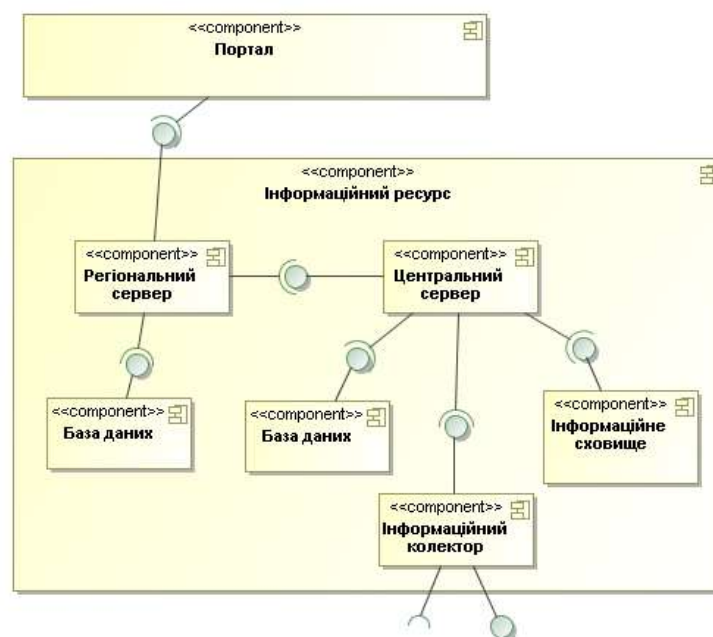


Рисунок 4 – Модель інформаційного ресурсу ДПА. Діаграма компонентів у нотації UML.

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 точки зору ресурсів до такої регіональної системи відноситься регіональна база даних. Регіональна БД зберігає основну інформацію, необхідну для роботи системи в регіонах. Для більшості запитів звернення до центрального серверу не є необхідним.

Центральний сервер зберігає дані та інформацію, необхідні для роботи системи в цілому. До них відноситься центральна база даних, інформаційний колектор та інформаційне сховище. Інформаційне сховище призначено для збору 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 корисними для користувачів. Колектор забезпечує зберігання інформації ДПА та доступ до інформації в колекторах інших засобів державної влади.

Кожен з компонентів деталізується при подальшому моделюванні. На апаратному рівні як регіональний, так і центральний сервер включає сервер застосувань, сервер баз даних та проксі-сервер. Така система забезпечує необхідний рівень надійності та безпеки.

ВИСНОВКИ

Застосування прикладного системно аналізу та компонентного підходу до проектування інформаційних ресурсів ІКС дозволяє упорядкувати і суттєво спростити процес проектування інформаційного ресурсу, врахувати конкретні вимоги, провести оптимізацію структурного та динамічного представлення інформаційного ресурсу, закласти необхідні рішення у відповідності із іменованою специфікацією на інформаційний ресурс [24].

Застосування для проектування графічної мови моделювання UML та засобів автоматизації процесу проектування забезпечує автоматизовану генерацію програмного забезпечення інформаційного ресурсу та тестування на кожній ітерації та фазі розробки.

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

Специфікація вимог до ресурсу за допомогою визначення інтерфейсів дозволяє розділити процес розробки ресурсу на незалежні компоненти, що можуть бути спроектовані і розроблені окремо. Визначені вимоги дозволять обґрунтовано обрати засоби створення окремих компонентів ресурсу. Використання сучасних стандартів взаємодії забезпечує надійне поєднання розроблених окремо компонентів на стадіях реалізації та розгортання.

ЛІТЕРАТУРА

1. Масляно П.П. Технология информатизации корпоративных структур, ч.1 // Корпоративные системы. - 2003. - №1. - С. 17-19.
2. Масляно П.П. Технология информатизации корпоративных структур, ч.2 // Корпоративные системы. - 2003. - №4. - С.17-19.
3. Кватрани Т. Визуальное моделирование с помощью Rational Rose 2002 и UML.: Пер. С англ. – М.: Вильямс, 2003.
4. Масляно П.П., Майстренко О.С. Моделювання бізнес-процесів організаційної структури // Вісник східноукр. національного університету ім. В. Даля. - 2007р. - №5. - С. 136-141.
5. Масляно П.П., Ліссов П.М. Інформаційні ресурси та засоби їх створення // Вісник Східноукр. нац. унів-ту ім. Володимира Даля. – 2007р. – №5 (111) – С. 141-145.
6. Масляно П.П. Концепція інформатизації корпоративних структур // Наукові вісті НТУУ "КПІ" - 2003р. - №3. - С.510-525.
7. Ракитина Е.А., Пархоменко В.Л. Информатика и информационные системы в экономике: Учеб. пособие. Ч.1. Тамбов: Изд. тамб. гос. техн. ун-та - 2005. - 148с.
8. Закон України "Про телекомунікації" (від 18.112003р. № 1280-IV).
9. www.answers.com.
10. Черненко М., Слепцов С. Принципы классификации управленческих информационных систем // Корпоративные системы. – 2004г. – №1.
11. Закон України "Про Національну програму інформатизації".

12. Костенко Л.Й., Сорока М.Б. Библиотека інформаційного суспільства // Бібл. вісн. — 2002. — №3. — С. 33-38.
13. Маслянюк П.П., Лісов П.М. Інформаційно-комунікаційні системи та технології обробки інформаційних ресурсів // Вісник КУЕІТУ «Нові технології» - 2007р. - №1-2 (15-16) - С. 20.
14. Маслянюк П.П., Лісов П.М. Проблеми і технології продукування інформаційних ресурсів // Матеріали Всеукр. наук.-практ. конф. «Сучасні тенденції розвитку інформаційних технологій в ...» 11-13 грудня 2006р., м. Луганськ, - С.184-189.
15. http://ru.wikipedia.org/wiki/Список_файловых_систем.
16. Codd, E. Derivability, Redundancy, and Consistency of Relations Stored in Large Data Banks // Communications of the ACM - 1970.
17. Сиротюк О. Особенности проектирования современных баз данных / www.computerworld.com.ua.
18. Inmon, W.H. Tech Topic: What is a Data Warehouse? // Prism Solutions. - 1995. - Vol. 1.
19. Маслянюк П.П., Лісов П.М. Системне проектування інформаційних ресурсів // Матеріали X Міжн. наук.-техн. конф. «Системний аналіз та інформаційні технології» - 20-24 травня 2008р., м.Київ - 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.
21. <http://www.worldcat.org/>.
22. Online Computer Library Center, <http://www.oclc.org>.
23. Tansley R. Building a Distributed, Standards-based Repository Federation, The China Digital Museum Project // D-Lib Magazine. - 2006. - vol.12, num. 7/8.
24. Маслянюк П.П., Лісов П.М. Дослідження та розробка підсистеми автоматизації проектування інформаційних ресурсів організаційних систем // Матеріали II Всеукр. наук.-практ. конф. «Сучасні тенденції розвитку інформаційних технологій ...», 8-10 квітня 2008 р., м. Луганськ - С.57-59.
25. Маслянюк П.П. Концепція управління безпекою інформації в корпоративних структурах // Ювілейна науково-технічна конференція „Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні”, Україна, Київ, 9-11 червня 1998.
26. Маслянюк П.П. Системне проектування процесів інформатизації // Науові Вісті НТУУ «КПІ» - 2008р. - №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.