

УДК 004.9

Д. С. ЗАМЯТІН, О. С. КЕБКАЛО, Л. І. ЛУКАШЕВСЬКА, А. Ю. МИХАЙЛЮК, О. С. МИХАЙЛЮК,
А. В. ПЕТРАШЕНКО, В. П. ТАРАСЕНКО

НТУУ "Київський політехнічний інститут", М. Київ

ОРГАНІЗАЦІЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ КОРПОРАТИВНОГО ДОКУМЕНТО-ОРІЄНТОВАНОГО СХОВИЩА ДАНИХ ОСВІТНЬОЇ УСТАНОВИ

Анотація. Пропонується концепція організації універсального програмного середовища, призначеного для формування відомчого документо-орієнтованого сховища даних навчального закладу з подальшою можливістю індивідуального конструювання засобів бізнес-логіки, загалом характерної для освітньої діяльності. В основу концепції ховища слабоструктурованих даних покладена логічна модель «об'єкт-атрибут-значення», яка дозволяє не лише зберігати електронні документи з різними наборами атрибутів, але й забезпечувати можливість оперативної модифікації цих наборів.

Ключові слова: інформаційні ресурси навчального закладу, корпоративне сховище даних навчального закладу, документо-орієнтоване сховище даних, сховище слабоструктурованих даних.

Аннотация. Предлагается концепция организации универсальной программной среды, предназначенной для формирования ведомственного документо-ориентированного хранилища данных учебного заведения с возможностью индивидуального конструирования средств бизнес-логики, характерной для образовательной деятельности. В основу концепции хранилища слабоструктурированных данных положена логическая модель «объект-атрибут-значение», которая позволяет не только хранить электронные документы с разными наборами атрибутов, но и обеспечивать возможность оперативной модификации этих наборов.

Ключевые слова: информационные ресурсы учебного заведения, корпоративное хранилище данных учебного заведения, документо-ориентированное хранилище данных, хранилище слабоструктурированных данных.

Abstract. A conception to organizing universal programming environment, intended to form the departmental document-oriented data warehouse of the educational institution, is offered in the article with possibility of individual design of business logic that is generally characteristic for educational activities. The concept of specialized warehouse of semistructured data is based on the logical model "object-attribute-value", which allows not only to store electronic documents with different sets of attributes, but also to provide the possibility of operative modification of these sets.

Key words: information resources of the educational institution, corporate data warehouse of the educational institution, document-oriented data warehouse, warehouse of semistructured data.

Вступ

Однією з характерних рис освітньої діяльності в суспільстві знань є виражена тенденція до використання переважно електронної форми подання навчальної інформації [1, 2]. Очевидно, це ставить перед системою освіти взагалі, і перед окремими навчальними закладами зокрема, ряд нових завдань, які потребують негайного вирішення. До числа таких завдань безумовно належить формування та упорядкування власного електронного інформаційного ресурсу (ІР) [3]. Частково ця задача розв'язується за рахунок використання специфічного програмного забезпечення (ПЗ) освітнього призначення [4]. Так, у системах дистанційного навчання, електронних підручниках, електронних бібліотеках тощо звичайно передбачають вбудований інструментарій для підтримки власного ІР. Однак, використання широкого переліку спеціалізованих систем породжує загалом негативне, з точки зору упорядкування інформації, явище технологічної та логічної гетерогенності ресурсу [5, 6]. Крім того, існує надзвичайно великий прошарок інформаційних об'єктів, що залишаються поза сферою впливу традиційних освітніх та офісних комп'ютерних систем [3, 7]. До числа таких об'єктів можна віднести електронні версії учнівських (студентських) робіт, різноманітні презентації, методичні матеріали, наукові розробки, будь-які електронні документи (Д), які з тих чи інших причин не вписуються у технологічний цикл традиційних систем електронного документообігу, і т.п. Ці дані формують окремий, також гетерогенний, інформаційний простір, який є надзвичайно важливим в плані використання у освітній та науковій діяльності учбового закладу, але не забезпечений конкретними ПЗ. Гетерогенність простору обумовлюється не тільки різноманітністю типів об'єктів, а й розмаїттям самих об'єктів одного типу. Об'єкти, що розглядаються, є слабоструктурованими, оскільки їх внутрішня організація залежить від багатьох факторів (тип Д, авторство, час створення тощо). Слід визнати, що перелік таких даних інтенсивно змінюється (в основному, збільшується), у зв'язку з чим використання для їх упорядкування деякої бази з наперед заданою сталою схемою даних не є реальним. Підтримка ж у повній мірі власного динамічного електронного ІР, передусім текстового, засобами стандартних систем керування базами даних сьогодні не під силу пересічним освітнім закладам, принаймні вітчизняним, зокрема з економічних причин [8].

Постановка задачі

Тому в контексті глобальної інформатизації освіти актуальною задачею є розробка та впровадження спеціалізованих, орієнтованих на використання в навчальних закладах, сховищ слабоструктурованих даних [9-12]. Далі пропонується підхід і концепція організації універсального програмного середовища, призначеного для формування відомчого документо-орієнтованого сховища даних навчального закладу, на базі якого існує можливість індивідуального конструювання засобів бізнес-логіки, характерної для різних ланок освітньої та наукової діяльності.

Загальна концепція спеціалізованого документо-орієнтованого сховища даних

В основу концепції спеціалізованого сховища слабоструктурованих даних покладена логічна модель «об'єкт-атрибут-значення», в якій кожна сутність або об'єкт подається певною сукупністю характеристик або атрибутів. Ця модель дозволяє не лише зберігати Д з різними наборами атрибутів, але й підтримувати динамічність цих наборів. Тому інформаційна база системи дозволяє накопичувати та обробляти сутності довільної структури (з довільною кількістю і типами атрибутів) [13].

Окремий ІР у межах сховища розглядається як підмножина документів, де окремий Д — довільний набір атрибутів і їх значень.

Розмежування прав доступу для роботи з ресурсами здійснюється на основі шаблонів. Шаблон — це набір умов для значень атрибутів Д, який дозволяє здійснити відбір потрібного сегменту ресурсів. Завдяки шаблонам групі користувачів можна призначити права в одному режимі роботи (у даному випадку і надалі маються на увазі режими роботи з Д) з одним набором Д, а у другому режимі — з іншим набором Д (певного ІР). Для ПЗ сховища даних (надалі - системи) визначено рівні доступу до ІР (гість, модератор тощо). Кожному рівню доступу відповідає набір режимів роботи, кожному з яких поставлено у відповідність шаблон. Можна призначити різні рівні доступу для різних ІР. Завдяки включенню користувача у довільний набір груп можна вибрати і призначити будь-яку комбінацію рівнів доступу до різних ІР. Відповідність Д певному шаблону визначається динамічно при роботі у кожному з режимів. Тобто, при зміні Д можливе включення або виключення його доступності у будь-якому з режимів для групи користувачів з певним рівнем доступу.

Для користувачів визначено наступні режими функціонування системи:

- додавання і редагування Д;
- перегляду Д;
- адаптивної каталогізації;
- адміністрування.

Режим додавання і редагування Д передбачає:

- створення, редагування, видалення, групування, розгрупування атрибутів;
- додавання, редагування Д (включення і вилучення з них атрибутів, встановлення і редагування значень, збереження змін).

Режим перегляду Д:

- перегляд повного набору атрибутів Д і їх значень;
- перегляд та друк Д у вигляді однієї з форм всіх шаблонів, яким відповідає Д;
- перегляд, друк і завантаження файлів, пов'язаних із атрибутами відповідного типу;
- модерування, закриття, перехід до редагування, вилучення Д;
- перегляд Д, з яким даний пов'язаний асоціативним зв'язком.

Режим адаптивної каталогізації:

- вибір Д для подальшого перегляду;
 - пошук за значеннями і вмістом атрибутів;
 - перегляд звітів обробки Д;
 - систематизація, друк Д.
- Режим адміністрування:
- створення, редагування, видалення інформаційних об'єктів системи (атрибути, шаблони, профілі користувачів та ін.);
 - встановлення порядку відображення атрибутів Д при перегляді і редагуванні набору атрибутів шаблону;
 - конструювання формул, конструювання форм перегляду Д і форм звітів.

Адміністрування організоване таким чином, що для кожного типу інформаційних об'єктів (атрибут, шаблон, форма тощо), які використовуються у системі, можна додавати дозволи на виконання дій над ними. Кожному користувачеві або групі користувачів можна надавати ці дозволи або позбавляти їх. Для груп є можливість надання і позбавлення дозволів і прав на режими роботи з сегментами ІР. Шляхом включення певної категорії користувачів у групу можна автоматично надати їм дозволи і права, які визначені для даної групи. При роботі у кожному з режимів будь-якому користувачеві надаються лише ті можливості, які доступні йому згідно його прав.

Архітектура програмного забезпечення сховища даних

До складу системи входять наступні підсистеми (ПС) (рис. 1):

інформаційна база (ПС збереження ІР), відповідає за збереження ІР, які використовуються іншими модулями системи;

ПС адміністрування, забезпечує підтримку сесій, налаштування шаблонів, розмежування прав у кожному з режимів для різних ІР на основі шаблонів;

ПС введення даних, відповідає за наповнення і оновлення інформаційної бази;

пошукова ПС, здійснює каталогізацію, реалізує операції пошуку і встановлення асоціативних зв'язків;

ПС обробки даних, реалізує аналітичну обробку текстів, обслуговує користувацькі режими обробки Д і статистичну обробку даних;

ПС візуалізації, забезпечує формування інтерактивного інтерфейсу для користувача.

Далі подано опис складових частин вказаних підсистем.

Інформаційна база включає наступні сегменти даних.

Сегмент даних про атрибути — властивості атрибутів (назва, тип, текст підказки, можливість (не)відображення підказки, значення за замовчанням, порядок у списку атрибутів при перегляді, параметр для задання групових атрибутів та ін.), з яких складаються Д.

Сегмент Д - Д, які можуть належати до різних їх категорій.

Сегмент освітніх ІР різного формату — файли, які можна пов'язувати зі значеннями атрибутів типу «Файл» Д (навчально-методичні матеріали, фото і т.д.) і які можуть бути довільного формату.

4) *Сегмент адміністративних даних* — інформація про користувачів (профілі користувачів): логін, пароль, прізвище, ім'я, e-mail, привілеї (активність, статус персоналу, статус суперкористувача, дозволи на виконання дій над інформаційними об'єктами), останній вхід, дата приєднання, входження у групи тощо. Статус суперкористувача визначає, що такий користувач має всі дозволи без їх точного зазначення. Тут також міститься інформація: про групи користувачів, для яких можна визначити права на виконання дій над інформаційними об'єктами системи (наприклад, на додавання, редагування і видалення атрибутів) та права на режими роботи з відповідними ІР (ресурс публікацій, ресурс педагогічного досвіду тощо), які визначаються умовами шаблонів; про підтримку шаблонів — умови

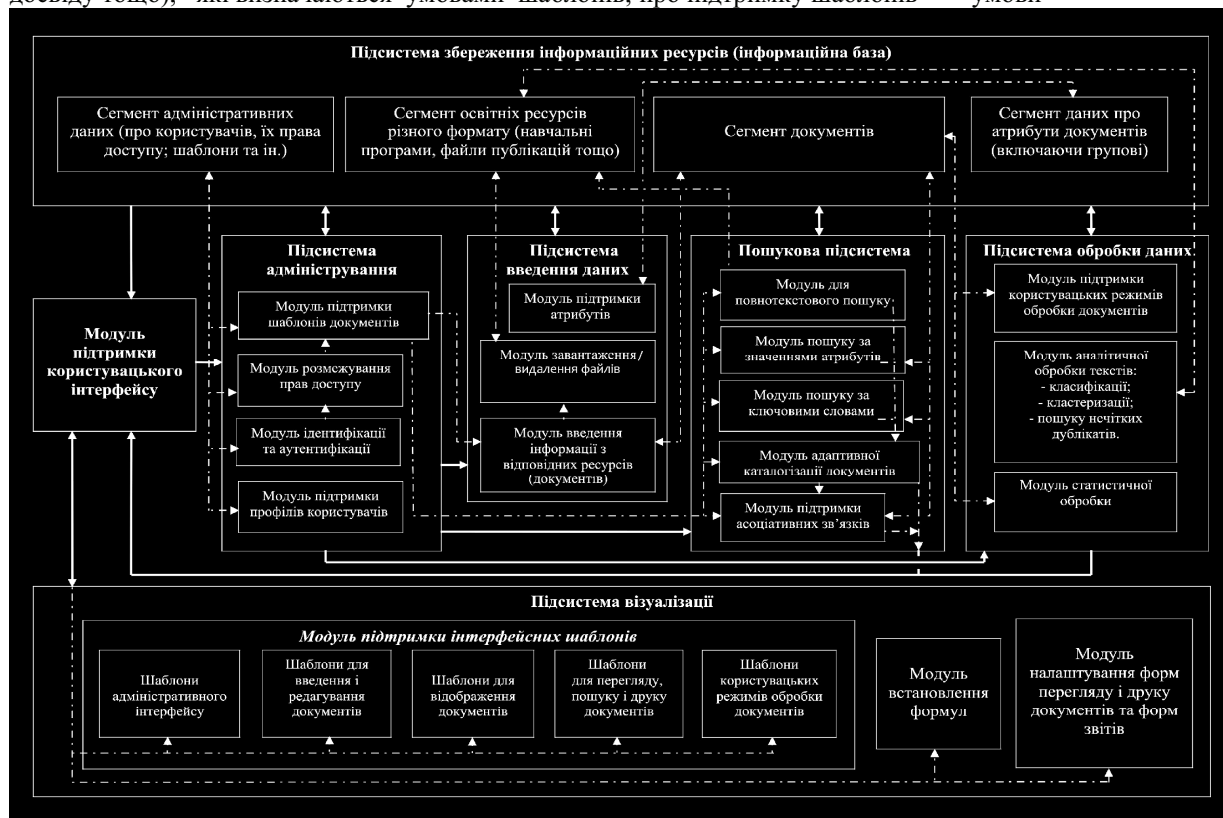


Рисунок 1 - Архітектура програмного забезпечення сховища даних навчального закладу

шаблонів, порядок відображення атрибутів для кожного шаблону; про параметри відображення форм перегляду Д та форм звітів, параметри звітів (формули). Для кожного шаблону або виду звіту можна створювати форми, у параметрах відображення яких задається інформація про координати розміщення і розміри блоків для атрибутів і їх значень (для форм Д) або для результатів звітів, а також для блоків міток.

ПС адміністрування має наступні складові.

Модуль підтримки профілів користувачів — дозволяє створювати, редагувати і видаляти профілі користувачів.

Модуль ідентифікації та аутентифікації — забезпечує можливість введення користувачем логіну і паролю та встановлення відповідності користувача пред'явленому ним ідентифікатору шляхом співставлення введених значень з інформацією, наявною у сегменті адміністративних даних.

Модуль розмежування прав доступу — у відповідності з правами доступу надає користувачеві можливість роботи у доступних режимах з доступними ІР системи, які визначаються умовами шаблонів. Для системи визначені наступні режими роботи з Д: *додавання, редагування, модерування, перегляд*. Режим модерування призначений для контролю за правильністю введених даних. Право на модерування надає користувачеві можливість обмежити доступність редагування певного набору Д для групи інших користувачів, після перевірки на коректність введених ними значень атрибутів. Таким чином, користувач системи, у обов'язки якого входить лише введення даних для певного сегменту ІР, може редагувати доступні йому Д доти, доки їх не перевірить і не зафіксує (промодерує) відповідальна особа.

Модуль підтримки шаблонів — дозволяє додавати, редагувати і видаляти шаблони і умови для них, здійснювати відбір Д, відповідних певному шаблону, а також визначати порядок відображення атрибутів для наборів Д, які відповідають певному шаблону. Для шаблону можна задавати умови не лише для значень атрибутів, які доступні для режиму редагування, але й атрибутів, які мають адміністративний характер («Модерування», «Власник»). Завдяки цьому забезпечується функціонування режиму модерування і фільтрація Д, створених певним користувачем, з метою надання йому прав лише на роботу із сегментом власних Д у визначеному режимі.

ПС введення даних включає наступні модулі.

Модуль підтримки атрибутів — дозволяє створювати, редагувати і вилучати атрибути з усіма їх властивостями, групувати і розгрупувати атрибути для користувачів, у яких є дозвіл на зміну інформаційних об'єктів даного типу. Аналогічні можливості надаються у підсистемі адміністрування, але на відміну від підсистеми введення даних з дещо іншою логікою діалогу і ширшими можливостями та зручнішими інструментами для відбору необхідних атрибутів (сортування за певними полями, пошук за назвою, фільтрація за типом).

Модуль введення Д — дозволяє вводити і редагувати Д для різних ІР. Конкретний користувач може вводити (редагувати) лише ті Д, які відповідають шаблону, визначеному для режиму введення (редагування) для груп, у які включений даний користувач. При введенні даних користувачеві пропонується обрати шаблон, після чого відображається певний набір атрибутів, який динамічно формується у базі даних у процесі перебору всіх існуючих Д. Таким чином, набір атрибутів шаблону — це сукупність всіх існуючих атрибутів збережених Д, котрі відповідають обраному шаблону. Додавання нового атрибуту в Д призводить до появи даного атрибуту у наборах атрибутів всіх шаблонів, яким відповідає вказаний Д. Якщо при зміні значення одного з атрибутів Д атрибут втрачає відповідність певному шаблону, то атрибути даного Д не враховуються при формуванні набору атрибутів даного шаблону. До Д можна включати атрибути, які уже присутні, для введення декількох значень для одного атрибуту.

Модуль завантаження/видалення файлів — дозволяє завантажувати з файлової системи на сервер і видаляти зі сховища освітні ІР або довідкові файли, які пов'язуються зі значеннями атрибутів типу «Файл», при введенні і редагуванні Д. У режимі перегляду модуль забезпечує завантаження з серверу необхідних файлів для детальнішого ознайомлення з їх змістом.

ПС пошуку містить наступні модулі.

Модуль адаптивної каталогізації — систематизує Д сховища, надає структуровану інформацію про Д і можливість її застосування для подальшого пошуку, забезпечує доступ до Д за атрибутами. Перегляд Д здійснюється через каталог, у якому відображається той набір Д, який доступний користувачу у цьому режимі згідно його прав. Будь-який з Д, доступних для перегляду, може мати довільний набір атрибутів. Користувачеві надається список всіх атрибутів, що присутні у доступному для перегляду наборі Д. При виборі атрибуту відображається список всіх його можливих значень у алфавітному порядку серед всіх Д набору, які можуть бути використані для подальшого пошуку. Список атрибутів змінюється відповідно до зміни набору Д.

Модуль для повнотекстового пошуку — дозволяє здійснити пошук серед файлів, які пов'язуються з Д. Даний модуль також має реалізувати пошук за ключовими словами у тексті файлів Д і використовувати порядок відображення результатів, який залежить не лише від частоти входження вказаних слів, але і від інших критеріїв, які підвищують релевантність пошуку, за рахунок впровадження засобів Text Mining.

Модуль пошуку за значеннями атрибутів — дозволяє відфільтрувати Д згідно заданих значень атрибутів. Даний вид пошуку може бути ієрархічним. Таким чином, реалізується операція перетину для множин Д з обраними значеннями атрибутів.

Модуль пошуку за ключовими словами — дозволяє знайти Д, у значеннях відповідних атрибутів яких містяться введені користувачем ключові слова (регістронезалежний пошук). Пошук за ключовими словами може здійснюватись для значень довільного набору атрибутів. Таким чином, будуть відібрані ті Д, які мають вказані користувачем атрибути, і у відповідних значеннях яких містяться введені ключові

слова. Пошук може бути комбінований (взаємодіючий). Можна також паралельно використовувати ці два види пошуку (за значеннями атрибутів і за ключовими словами), що забезпечує фільтрацію Д за значеннями і вмістом відповідних атрибутів.

Модуль підтримки асоціативних зв'язків — дозволяє знайти за допомогою засобів адаптивної каталогізації ті Д, які пов'язані з обраним, і встановити асоціативні зв'язки за рахунок атрибутів відповідного типу. Якщо у користувача немає прав на перегляд Д, які пов'язані асоціативним зв'язком з доступним, то відповідні атрибути, які встановлюють вказаний зв'язок, взагалі не відображаються.

ПС обробки даних включає наступні модулі.

Модуль підтримки користувацьких режимів обробки Д — забезпечує реалізацію необхідних користувачеві задач шляхом здійснення операцій (вибірка, фільтрація та ін.) над необхідними наборами Д. В основі лежить використання синтаксичного аналізатора формул і математичних виразів, завдяки чому забезпечується підтримка змінних вимог до проведення обробки інформації. При необхідності здійснення операцій над новоствореним сегментом Д з новими атрибутами відповідальна особа вводить формули з відповідними параметрами, які визначаються логікою обробки.

Модуль аналітичної обробки текстів забезпечує наступні процеси.

Класифікацію — групування вихідних понять, наявних у тексті, за будь-якими суттєвими ознаками і розташування їх у певному порядку, що відображає ступінь схожості.

Кластеризацію — логічне продовження класифікації, але класи набору даних заздалегідь невідомі. Призначена для розбиття сукупності об'єктів на однорідні групи.

Пошук нечітких дублікатів (для виявлення плагіату). Забезпечує визначення того, чи фрагменти даного тексту взяті із уже існуючих, тобто наявних у сегменті освітніх ІР інформаційної бази системи.

Модуль статистичної обробки забезпечує застосування засобів статистики (статистичних показників, критеріїв) для формування даних, необхідних для звітної документації. *ПС візуалізації* містить наступні складові частини.

Модуль підтримки інтерфейсних шаблонів — забезпечує генерацію веб-сторінок з необхідною інформацією і включає наступні групи шаблонів.

Шаблони адміністративного інтерфейсу призначені для привілейованих користувачів (адміністраторів). Забезпечують інтерфейс з елементами управління і навігацією, які дозволяють виконувати дії адміністративного характеру. Для кожного типу об'єктів існує форма редагування, яка дозволяє додавати, редагувати і видаляти конкретні записи і список всіх доступних у базі даних об'єктів. У формі редагування для кожного типу значень визначено відповідні елементи відображення (календар для типу «Дата», перемикач для булевого типу тощо). У списку об'єктів забезпечуються зручні візуальні засоби для здійснення пошуку, фільтрації, сортування змінюваних інформаційних об'єктів. При встановленні порядку для атрибутів шаблону здійснюється динамічне відображення набору атрибутів даного шаблону у порядку, який був збережений у базу. Для зміни цього порядку використовується технологія drag-and-drop (перетягування за допомогою миші).

Шаблони для введення і редагування Д. Забезпечується відображення атрибутів у встановленому порядку у відповідному вікні з метою надання користувачеві можливості вибору і включення певного атрибуту в Д. Елементи інтерфейсу забезпечують зручний режим введення (редагування) Д за рахунок:

використання динамічного відображення атрибутів, які відповідають обраному шаблону;

спливаючих підказок для введення значень;

різних форм відображення елементів для введення значень атрибутів (перемикачі, прапорці, текстові поля, поля введення, випадаючі списки, поля типу "Файл", спливаючі календарики для введення значень типу "Дата");

діалогових вікон з адаптивною каталогізацією (для атрибутів типу "Асоціативний зв'язок") та ін.

Для кожної з форм відображення елементів введення, якщо значення для доданого в Д атрибуту не було встановлено, атрибут зберігається у Д і доступний при наступному редагуванні без додаткових включень.

Шаблони для відображення Д. Атрибути Д відображаються у встановленому модулем підтримки шаблонів порядку. Для певних форматів завантажені освітні ІР можна переглядати у відповідному вікні і здійснювати друк. Всі файли можна завантажувати з серверу, користуючись посиланням. У даному шаблоні перегляду можна відобразити Д у вигляді будь-якої форми інших шаблонів, яким відповідає даний Д, а також роздрукувати форму. Для атрибутів типу «Асоціативний зв'язок» значення відображаються у вигляді гіперпосилання, за допомогою якого можна перейти до перегляду відповідного Д.

Шаблони для перегляду, пошуку і друку Д. Наявна посторінкова навігація доступних для перегляду Д. Пошук за ключовими словами полегшується за рахунок використання спливаючих підказок, а пошук за значеннями атрибутів — за рахунок динамічного відображення значень у алфавітному порядку. Для друку обраних Д можна обирати атрибути, значення яких відображені у таблиці для виведення інформа-