

Олексій Костишин, Володимир Широков

Трирівнева система цифрових бібліотек з технологією «клієнт—сервер»

Розглянуто принципи побудови спеціалізованих цифрових бібліотек із застосуванням технології «клієнт—сервер». Досліджено архітектуру подібних систем, призначених для роботи у мережесередовищах інтранет/Інтернет. Вивчено принципи побудови підсистеми клієнта та серверних підсистем — доступу, застосувань, архівації і зберігання та запропоновано набір компонентів програмного забезпечення, необхідний для реалізації платформонезалежних систем цифрових бібліотек, що працюють у мережевому оточенні.

Розвиток інформаційних технологій у мережі Інтернет не міг не вплинути на розвиток традиційних сховищ даних — бібліотек. Насамперед зміни торкнулися систем каталогізації, і традиційні паперові каталоги провідних бібліотек світу були переведені в електронну форму. Більш того, бібліотеки, що об'єдналися під егідою OCLC (Online Computer Library Centre), створили найбільший у світі електронний каталог (понад 40 млн бібліографічних описів).

У середині 90-х років минулого сторіччя ряд дослідницьких організацій США запропонував проект цифрових бібліотек — інформаційних систем, що поєднують цифрові бібліографічні каталоги (системи OPAC) і, власне, первинні інформаційні ресурси (книги, журнали тощо у цифровій формі). Проект одержав назву Ініціативи цифрових бібліотек (Digital Library Initiative — DLI: <http://www.dlib.org>); його засновниками стали такі організації, як DARPA (Defence Advanced Research Project Agency — Агентство передових дослідницьких проектів Міністерства оборони США), NASA і NSF (National Scientific Foundation — Національний фонд наукових досліджень США). Цікаво, що серед цих організацій немає жодної бібліотеки у її традиційному розумінні.

У межах проекту DLI було розроблено, розвинено та впроваджено стандарти електронних бібліографічних описів і форматів збере-

ження цифрових гетерогенних даних (текстів, графіки, зображень, звукових та відеоматеріалів). На основі цих стандартів почалися експерименти з проектування цифрових бібліотек та їх інтеграції в глобальну мережу Інтернет. Результатом експериментів стала реалізація цифрових бібліотек у провідних науково-дослідних і навчальних закладах США: Масачусетський інститут технологій, Каліфорнійський університет (Берклі) та ін. Для прискорення доступу користувачів до періодичних видань деякі редакції журналів та газет стали розмішувати цифрові копії своїх видань у мережі Інтернет.

Саме поняття терміна «цифрова бібліотека» («digital library») базується на аналогії зі звичайною бібліотекою і досить адекватно відображає загальні функції автоматизованих інформаційних систем подібного класу. В електронній бібліотеці присутні аналоги всіх основних елементів, характерних для традиційної бібліотеки, а саме:

- система каталогів, що відіграє роль вхідної підсистеми бібліотеки та є її інформаційно-пошуковим апаратом;
- «сховище» об'єктів, у якому містяться в електронній формі книги, журнали, газети, інша інформаційна продукція;
- комплекс сервісних і технологічних засобів (реєстрація й обслуговування користувачів, каталогізація, систематизація, облік, статистика, аналіз і зміни у фондах тощо).

Специфіка електронної бібліотеки порівняно з іншими типами інформаційних систем, найближчими з яких за своїми функціями і структурою є повнотекстові бази даних та електронні документосховища, полягає в тому, що в електронній бібліотеці набагато повніше, більш структуровано і стандартизовано представлені всі основні компоненти й інформаційно-лінгвістичне забезпечення, до яких необхідно віднести:

- пошуковий апарат (електронний каталог) системи формується на основі узагальненого поняття бібліографічного опису об'єкта збереження, що є предметом доволі роз-

Костишин Олексій Максимович, к. фіз.-мат. н., с. н. с. Українського мовно-інформаційного фонду НАН України,

Широков Володимир Анатолійович, д. т. н., директор Українського мовно-інформаційного фонду НАН України.

галуженої системи уніфікації і стандартизації (ISO-2709, MARC, USMARC, UNIMARC тощо);

- система узагальнених об'єктів збереження ґрунтується на сучасних форматах представлення гетерогенної інформації і відповідному інформаційному, лінгвістичному і програмному забезпеченні Multy та Hypermedia систем, що дозволяє вводити, зберігати й обробляти в цифрових бібліотеках дані будь-якого типу (зокрема текстову, графічну, аудіо-, відеоінформацію тощо). Останнє зближує електронні бібліотеки з електронними системами керування документами — Electronic Document Management Systems;
- сервісно-технологічна частина електронної бібліотеки базується на досвіді формування й експлуатації традиційних «паперових» і т.зв. «гібридних» бібліотек (<http://www.ukoln.ac.uk/services/elib/projects>) з можливістю реалізації всіх доцільних у цьому випадку «бібліотечних» функцій, а також додаткових функцій, які враховують специфіку сучасних автоматизованих інформаційних систем.

Електронні бібліотеки, так само, як і традиційні, поділяються на універсальні й спеціалізовані. Для універсальної електронної бібліотеки характерні:

- найбільш повна реалізація стандарту бібліографічного опису, доповнення формату бібліографічного запису повнішим набором інформаційних елементів бібліографії на основі уніфікованих правил каталогізації;
- широке застосування стандартних елементів інформаційно-лінгвістичного забезпечення (класифікаторів, словників, тезаурусів і т.п.);
- розширена система форматів подання гетерогенної інформації в базі узагальнених об'єктів, що забезпечує надійне й ефективне представлення об'єктів електронної бібліотеки;
- широкий набір функцій системного сервісу, що не тільки не поступається, але й перевищує відповідний набір системного сервісу традиційних бібліотек.

Спеціалізовані електронні бібліотеки, як правило, відіграють роль складових частин великих інформаційних систем, до яких вони входять як підсистеми. З цього випливають певні вимоги до їхніх функцій і структури. У підсистемі електронного каталогу може застосовуватися обмежений формат бібліографічного опису (що не повинно негативно позначатися на його відповідності стандартам). Разом

з тим, тут можуть бути реалізовані додаткові функції, що впливають зі специфіки системи — інформаційно-статистичні, семантичні, лінгвістичні, когнітивні, прогнозні і т. ін.; засоби транспорту бібліографічної інформації з електронного сховища даних у видавничу підсистему і т. д.

Методи, а також інформаційні і програмні засоби, що застосовуються для представлення узагальнених об'єктів збереження в спеціалізованій електронній бібліотеці, мають орієнтуватися на специфіку згаданих об'єктів. Наприклад, в електронній бібліотеці технічної документації варто передбачити програмне забезпечення підтримки CAD-систем, що дає змогу користувачеві працювати з конструкторською документацією.

Щодо змісту поняття «електронна бібліотека», то варто застерегти відносно глибини її аналогії з традиційними бібліотеками. Термінологічна і предметна близькість цих понять здатні привести до непорозумінь, оскільки насправді йдеться про об'єкти істотно різні за своєю природою. Електронні бібліотеки — це, насамперед, автоматизовані інформаційні системи, що проектуються відповідно до законів та стандартів проектування комп'ютерних систем, а технології формування, ведення і застосування електронних бібліотек зазвичай не дуже нагадують стандартні бібліотечні технології.

В Українському мовно-інформаційному фонді (УМІФ) НАН України з 1991 р. ведеться робота над створенням спеціалізованої електронної бібліотеки, специфіка якої визначена основним напрямом досліджень робіт УМІФ — комп'ютерною лінгвістикою. Зазначена бібліотека є складовою частиною більш широкої та інтегрованої системи — лексикографічного процесора, призначеного для проведення лінгвістичних досліджень та розробок. Саме ця специфіка визначає ті системотехнічні засади, на яких вона повинна бути побудована, а також режим її експлуатації.

З означених міркувань електронна бібліотека побудована за схемою, що на сьогодні є стандартом *de facto* в архітектурі розподілених корпоративних сховищ інформації, а саме за багаторівневою (*multi-tier*) схемою «клієнт—сервер». Подібна схема дозволяє здійснити системотехнічну декомпозицію електронної бібліотеки на дві базові складові частини — підсистему доступу й підсистему збереження.

Підсистема доступу, яка забезпечує користувачам доступ до необхідних ресурсів, у свою чергу, поділяється на дві основні частини, а саме «клієнт» (локальний або віддалений), у якій формується запит і куди спрямовуються результати його виконання, і «сервер застосу-

вань», який забезпечує обробку запиту, взаємодію з підсистемою збереження даних для їх одержання й передання результатів запиту клієнтові, який його сформував.

Зазначену підсистему на сьогодні реалізовано в середовищі локальної мережі, побудованої за *intranet*-технологією під керуванням операційної системи Windows 2000 Advanced Server (серверна частина) та Windows ME/2000 Professional (частина клієнта). У ролі власне клієнта може виступати будь-який NCSA/Mosaic-сумісний Web-броузер — Microsoft Internet Explorer, Netscape Navigator, Opera і подібні.

Сервер застосувань виконано на базі HTTP-демона Apache (<http://httpd.apache.org>) з під'єднаними до нього модулями сучасних мов обробки сценаріїв — ModPerl (вбудований інтерпретатор Perl <http://www.ActiveState.org>) та ModPHP (вбудований інтерпретатор PHP <http://www.php.net>). Саме зазначені модулі реалізують сервер застосувань, виконуючи аналіз запитів клієнта *на стороні сервера* і утворюючи інтерфейс обміну інформацією (запити та їх результати) з підсистемою збереження. Слід зауважити, що використання HTTP-демона Apache дозволяє будувати платформонезалежні застосування — реалізації сервера Apache існують практично для всіх операційних систем (будь то Windows або довільний клон поширеної операційної системи UNIX).

Для реалізації підсистеми збереження в електронній бібліотеці, як і в широко поширених на сьогодні інформаційних сховищах, використано систему управління реляційними базами даних (СУРБД) на базі SQL сервера. Таке рішення, по-перше, зменшує обсяг інформації, яка циркулює у локальній мережі (від основної системи обробки та зберігання до клієнта передаються лише результати конкретного запиту, а не вся база даних цілком), по-друге, підвищує гнучкість та надійність системи зберігання, і, нарешті, дозволяє розділити технічні засоби, на яких реалізовано різні компоненти бібліотеки (сервер застосувань та сервер збереження).

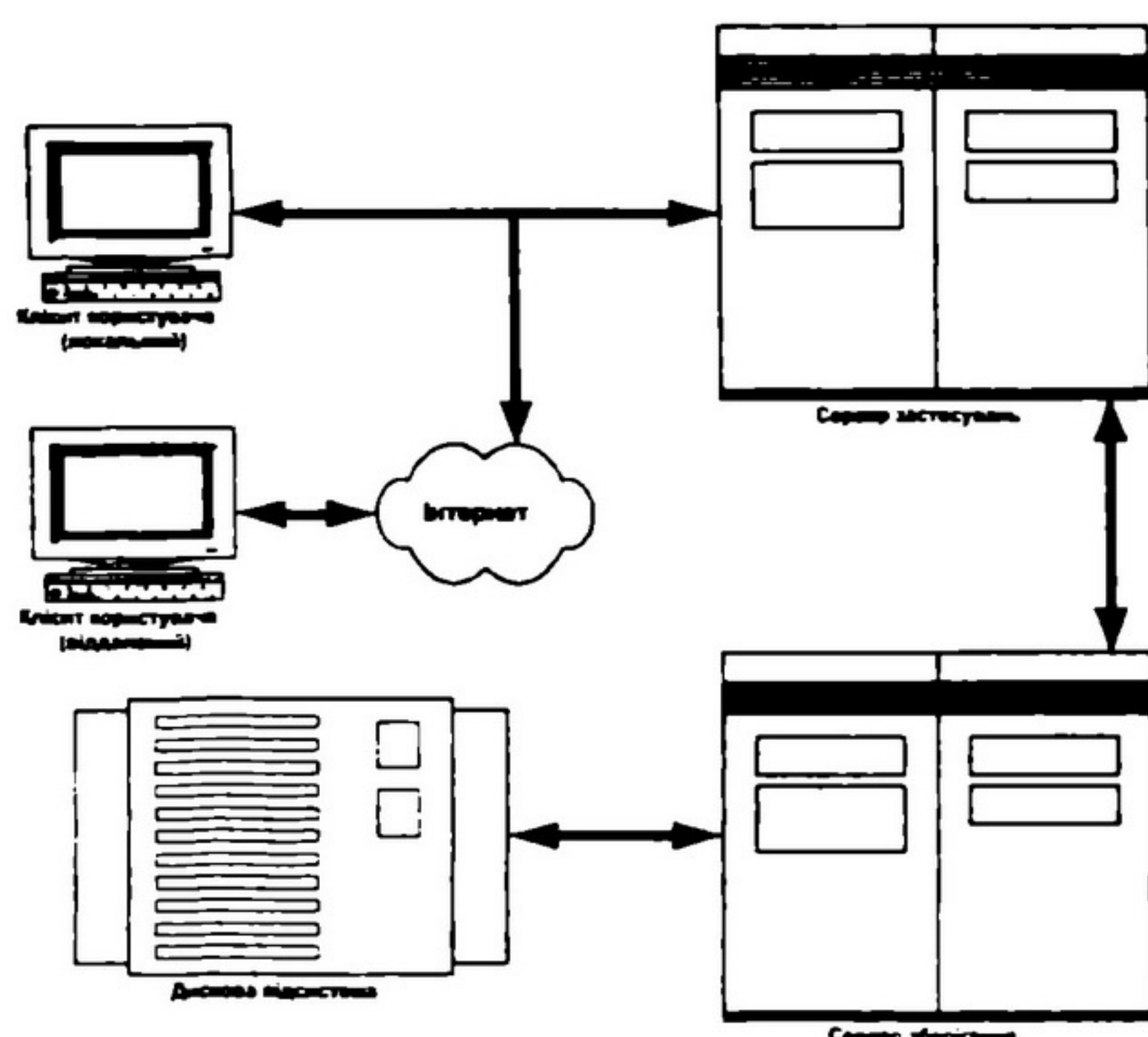
Відзначимо, що на сьогодні у пострадянських країнах в електронних бібліотечних системах досить поширена та використовується система СУБД CDS-ISIS або інші, створені на її основі (наприклад Irbis; створена у Державній науково-патентній бібліотеці Росії). Ця СУБД набула широкого поширення саме завдяки тому, що вона *ab initio* була орієнтована на використання у системах каталогізації бібліотек. Окрім того, розповсюджується вона безкоштовно. Але, як показав досвід побудови та експлуатації електронної бібліотеки в УМІФ, використання СУБД CDS-ISIS у сучасних інформаційних системах такого типу стає сумнівним. Це пов'язане з тим

фактом, що СУБД CDS-ISIS є нереляційною ієрархічною СУБД, яка підтримує досить обмежений набір типів даних (числові та символні), вона не підтримує гнучкої мови запитів (типу SQL) та не може працювати з сучасними системами кодування для підтримки багатомовних документів (стандарт UNICODE, <http://www.unicode.org>). Окрім того, СУБД CDS-ISIS не працює (вона й не була призначена для цього) з великими об'ємами даних бінарного типу (потік байтів, або BLOB), а саме у цьому форматі зберігається нетекстова інформація — малюнки, фотографії тощо.

При виборі програмного забезпечення для сервера зберігання нами було розглянуто два альтернативні варіанти. Перший з них — призначений для систем, що працюють у середовищі Windows 2000 Advanced Server — Microsoft SQL Server 2000. Головними причинами для вибору Microsoft SQL Server стали його поширеність у світі, повна підтримка розширеного стандарту мови ANSI SQL-92 (Transact-SQL), можливість роботи у локальних та глобальних мережах, підтримка надвеликих (до 64 Тбайт) баз даних, багатопроцесорних систем та багатомовних документів (стандарт UNICODE). Основним недоліком є сильна платформозалежність. Дане програмне забезпечення працює лише у середовищі Windows та на апаратних платформах x86. Окрім того, вартість Microsoft SQL Server 2000 досить велика. Наприклад, повний комплект програмного забезпечення з ліцензією на 5 осіб коштує близько \$1500.

Другий можливий варіант — SQL сервер MySQL шведської фірми ТсХ АВ (<http://www.mysql.org/download/>). При тих же причинах, що і у Microsoft SQL Server, система MySQL придатна для роботи на будь-яких сучасних програмних платформах UNIX, Linux, Solaris та ін. Доречно буде зауважити, що СУРБД MySQL, як і сервер Apache, стандартно включені до дистрибуції різних версій операційної системи Linux. Відзначимо, що це програмне забезпечення розповсюджується безкоштовно (згідно з GNU Public License).

Структурна схема спеціалізованої електронної бібліотеки УМІФ наведена на рисунку. Сьогодні вона використовується у проекті створення національної словникової бази згідно з Указом Президента України № 967 від 07.08.1999 року. Будучи невід'ємною частиною лексикографічного процесора, електронна бібліотека постачає інформацію для автоматизованої лексичної картотеки, яка необхідна при укладанні словників. Завдяки тісній інтеграції таких компонентів лексикографічного процесора, як електронна бібліотека та автоматизована лексична картотека, відпадає необхідність зберігати лек-



Структурна схема електронної бібліотеки, побудованої за технологією «клієнт—сервер»

сичні картки у явному вигляді — для будь-якого слова з реєстру нового словника вони є віртуальними об'єктами і генеруються автоматично на час їх необхідності.

Таким чином можна відзначити:

- ◆ Електронна бібліотека описаної архітектури може зберігати не тільки бібліографічні описи книжок або інших об'єктів зберігання (зображень, звукових та відеоматеріалів), а й самі об'єкти.

- ◆ Завдяки клієнт-серверній технології значно зменшене завантаження мережі, що є особливо важливою умовою при використанні такої системи у мережі Інтернет.
- ◆ Завдяки своїй платформи-незалежній архітектурі подібна електронна бібліотека природним чином з мінімальними витратами інтегрується як у мережу Інтернет, так і у будь-які корпоративні інформаційні середовища.

І, наостанок, слід додати, що в УМІФ проводяться дослідження з інтелектуалізації систем пошуку та аналізу в інтегрованих інформаційних системах, зокрема в системах електронних бібліотек та медіатек, причому інтерфейс запитів у таких системах пошуку та вилучення інформації може бути сформований на природній мові. Як показали дослідження, для реалізації такого інтерфейсу систем збереження інформації, реалізованих на традиційних СУРБД з фіксованими типами даних, буде недостатньо. Перспективним напрямом для реалізації подібних систем є розвиток так званих постріляційних, або об'єктно-реляційних систем управління базами даних та знань (наприклад постріляційна система Cache, <http://www.intersystems.com>), де потужність реляційних систем збереження інформації об'єднується з потужністю апарата об'єктно-орієнтованого програмування.

Костышин А., Широков В. Трехуровневая система цифровых библиотек с технологией «клиент—сервер»

Рассмотрены принципы построения специализированных цифровых библиотек с применением технологии «клиент—сервер». Изучена архитектура подобных систем, предназначенных для работы в сетевых средах интранет/Интернет. Подробно рассмотрены принципы построения клиентской подсистемы и серверных подсистем – доступа, приложений, архивации и хранения. Изучен набор компонентов программного обеспечения, необходимый для реализации платформонезависимых систем цифровых библиотек, функционирующих в сетевом окружении.

Kostyshyn O., Shyrovkov V. Three-leveled System of Digital Libraries with «Client-Server» Technology

Principles of constructing of specialized digital libraries with «client-server» technology are discussed. Architecture of semaltenious systems for operating in net media intranet/Internet was investigated. Principles for construction of client subsystems and server subsystems – access, using, archive and preservation were investigated. It was proposed a set of components for soft ware necessary for realization of platformed independend system of digital libraries operating in network media.

Kostyschyn O., Schyrovkov V. Dreiebenes System der Zifferbibliotheken mit Technologie «Kunde-Server»

Die Autoren untersuchen Aufbauprinzipien spezialisierten Zifferbibliotheken mit Anwendung der Technologie «Kunde-Server» und die Architektur ähnlicher Systemen, die man bei der Arbeit des Netzmediums Intranet/Internet verwendet. Es wurden auch die Aufbauprinzipien des Untersystems des Kunden und Serversystemen – Zugriff, Anwendung, Archivalien und die Erhaltung erforscht und Komplet der Software für Realisierung der dreiebenen Systemen der Zifferbibliotheken vorgeschlagen.

Kostychyn O., Chyrovkov V. Le système à trois niveaux des bibliothèques numériques utilisant la technologie «client – serveur»

On examine les principes de la création des bibliothèques numériques spécialisées avec recours à la technologie «client – serveur». L'article évoque l'architecture de systèmes semblables destinés au travail dans les environnement de réseau d'Intranet/Internet. On a étudié les principes de la création du sous-système de client et des sous-systèmes de serveur-access, mise en œuvre, archivation et conservation et il a proposé les composantes du génie logiciel en vue de la réalisation des systèmes de programme indépendants des bibliothèques numériques fonctionnés en entourage de réseau.