

Катерина Лобузінa,

зав. відділу НБУВ, канд. техн. наук

Роль бібліотечних класифікацій в організації знань

Публікація присвячена використанню бібліотечних класифікацій для упорядкування та представлення електронних інформаційних ресурсів різного характеру. Розглянуто сучасні засоби організації лінгвістичної інформації класифікаційної схеми. Наведено результати впровадження в Національній бібліотеці України імені В. І. Вернадського сучасних інструментів підтримки бібліотечної класифікації та удосконалення тематичного пошуку на їх основі.

К л ю ч о в і с л о в а: бібліотечні класифікації, семантична мережа, предметний пошук, тематична Інтернет-навігація.

The publication is dedicated to the use of library classifications for organizing and presenting electronic information resources of different nature. The modern means of organization of linguistic information classification scheme are considered. The results of implementation in The Vernadsky National Library of Ukraine modern library classification tools to support and improve contextual search based on it.

К e y w o r d s: library classification, semantic network, subject search, content Internet navigation.

Конвергенція (взаємопроникнення) бібліотечних та інформаційних технологій відбувається не тільки з боку освоєння бібліотеками нового інформаційного середовища, а й усвідомлення спеціалістами з інформаційних технологій того факту, що засоби організації знань необхідно шукати саме у бібліотеках. Про наявність таких інформаційних процесів свідчать, зокрема, концепції «системи організації знань» та «Семантичний Веб», що виникли у відповідь на недосконалість машинних методів пошуку інформації.

Ще у 1998 р. учасники конференції Association for Computing Machinery (ACM) (Міжнародна асоціація обчислювальної техніки) в Пітсбурзі [14], підбиваючи підсумки своєї роботи, вказали на необхідність розроблення та підтримки систем організації знань (СОЗ) (knowledge organization systems (KOS)), котрі позначають усі типи схем, що сприяють організації, керуванню, представленню, розкриттю та пошуку інформації (вони включають авторитетні файли контрольованої підтримки ключової інформації:

класифікаційні схеми, тезауруси, семантичні мережі та онтології). Згадані системи мають потенціал для опису, представлення, розкриття та пошуку гетерогенних інформаційних ресурсів Інтернету та створення єдиної інфраструктури розкриття знань. Вони призначені для організації пошуку інформації та обслуговування колекцій електронних документів. СОЗ виконують роль моста між користувачем, який веде пошук інформації, та матеріалами колекції, вони забезпечують можливість знайти необхідні користувачеві інформаційні джерела без попереднього знання про їх існування.

Семантичний веб (Semantic WEB (SW)) – концепція, розроблена Тімом Бернерсом-Лі (Tim Berners-Lee) [10] винахідником World Wide Web (WWW), URIs, HTTP та HTML. Його дослідник визначив як розширення (надбудову) існуючої Всесвітньої павутини з метою представлення інформації у вигляді, що дасть змогу краще розуміти її людям та комп'ютерам. Семантичний веб розглядався Т. Бернерсом-Лі в контексті розвитку Всесвітньої па-

вутини, як WWW другого покоління, що зорієнтований на автоматизовану інтерпретацію та оброблення інформаційних ресурсів. Він підтримується Міжнародним консорціумом «W3C» (The World Wide Web Consortium) [20]. Першим серйозним та популярним проектом, заснованим на принципах семантичної павутини, став проект метаданих «Дублінського ядра» (Dublin Core Metadata Initiative, DCMI) [13].

Важливу роль у концепції метаданих DCMI відіграють контрольовані словники – значення для відповідного поля метаописів обираються із фіксованої множини слів або кодів, котра обмежена набором ретельно підібраних пошукових термінів [1]. Це покращує можливості автоматичного оброблення метаописів, а також підвищує якість результатів пошуку. Крім уточнення самих елементів кваліфікований формат DC дає змогу уточнювати типи значень елементів. Це робиться через схеми кодування. Останні або уточнюють словник значень, з яких можна вибрати потрібні, або фіксують певний формат, у якому має бути представлений рядок-значення. Для елемента метаопису DC : Subject – предмет (тема), DC рекомендує схеми кодування, що ґрунтуються на відомих бібліотечно-бібліографічних класифікаційних системах: DDC (Dewey Decimal Classification – Десяткова класифікація Дьюї, <http://www.oclc.org/dewey/>); LCC (Library of Congress Classification – Класифікація Бібліотеки Конгресу, <http://www.loc.gov/catdir/cpso/lcco/>); LCSH (Library of Congress Subject Headings – Предметні рубрики Бібліотеки Конгресу, <http://www.loc.gov/cds/lcsh.html>); MeSH (Medical Subject Headings – Предметні рубрики в галузі медицини, <http://www.nlm.nih.gov/mesh/>); UDC (Universal Decimal Classification – Універсальна десяткова класифікація, <http://www.udcc.org/>).

Створення ефективних інструментів тематичного пошуку є складною, тривалою та коштовною справою, тому реаліям сьогодення більше відповідає наслідування та удосконалення попередніх досягнень цієї галузі бібліотечно-інформаційної діяльності. Така спадкоємність у технології аналітичного опрацювання електронних документів зумовлена тим, що вони представляють ті ж знання, що й традиційні інформаційні джерела – книги, журнали, реферативні видання тощо – тільки у більш зручному, загальнодоступному, гнучкому вигляді [2].

Бібліотечна спільнота накопичила значний досвід зі створення та підтримки класифікаційних схем, який зберігає свою актуальність і для нової технології. Представити класифікаційні схеми на основі нової технології прагнуть усі провідні бібліотечно-інформаційні центри. Сьогодні розгорнуто міжнародні та

національні програми з розкриття та каталогізації ресурсів Інтернету, котрі передбачають використання традиційних бібліотечних класифікацій [17].

На підставі результатів дослідження використання універсальних бібліотечних класифікацій сучасними інформаційними сервісами можна зробити висновок: розвиток нової інформаційної технології електронної бібліотеки відновив інтерес до бібліотечних класифікацій, вони відіграють суттєву роль у представленні, упорядкуванні та розкритті інформаційних ресурсів електронних бібліотек, Інтернету. Аналіз використання традиційних бібліотечних універсальних класифікаційних схем в Інтернеті (17 інформаційних служб) [3] та електронних бібліотеках свідчить про те, що багатогалузеві класифікаційні схеми мають низку переваг при розкритті та представленні електронних інформаційних ресурсів. А саме, вони: більше ніж інші типи класифікацій відомі загалу користувачів; підтримуються великими бібліотечними центрами, мають добре розроблені методичні засади для підтримки в актуальному стані; охоплюють усі галузі знання та надають можливість обмежувати сферу пошуку, виділяти необхідну тематичну частину великої колекції. Все це створює передумови широкого застосування універсальних бібліотечних класифікацій для представлення та упорядкування електронних бібліотечно-інформаційних ресурсів різного характеру (реферативно-бібліографічні БД, електронні колекції, каталоги ресурсів Інтернету). Вирішального значення для організації доступу до електронних інформаційних ресурсів набуває створення програмного забезпечення для підтримки класифікаційних схем та авторитетних файлів класифікаційних даних.

У 1998 р. Керівною групою IFLA з розроблення формату авторитетних / нормативних записів було започатковано проект формату UNIMARC авторитетних записів для класифікаційних даних, який безпосередньо пов'язаний із форматом UNIMARC для обміну бібліографічними даними. Зусилля фахівців у цьому напрямі мали на меті представити таблиці різних класифікаційних систем в уніфікованому машиночитаному вигляді, автоматично прив'язати їх як пошуковий інструмент до бібліотечних баз даних, представлених у форматі UNIMARC. Формат UNIMARC для класифікаційних даних побудовано відповідно до структури формату USMARC для класифікаційних даних [18], розробленого у 1991 р. Бібліотекою Конгресу США. Наприклад, база даних програмного продукту Dewey for Windows організована відповідно до формату USMARC для класифікаційних даних [6; 12]. Відповідним чином

структурованi записи i еталонного файлу УДК, котрий розповсюджується Консорціумом УДК у середовищi CDS/ISIS. Кожний запис, що відповідає певному iндексу УДК, мiстить 30 полiв, згрупованих у двi категорiї. Першi двадцять полiв описовi i представляють фактичний змiст таблиць УДК, подальшi десять – надають адмiнiстративно-управлiнську iнформацiю [16].

Лiнгвiстичнi бази даних, якi використовуються у бiблiотечнiй практицi (словники предметних рубрик, класифiкатори i рубрикатори), забезпечують встановлення зв'язкiв мiж авторською термiнологiєю (документами), термiнологiєю iнформацiйно-пошукової системи (ПiС) та термiнологiєю споживачiв (професiйною термiнологiєю). Авторитетнi файли лiнгвiстичних даних є основним семантичним засобом бiблiотечних ПiС, вони створюють передумови реалiзацiї iнтелектуального тематичного пошуку, перетворюють бiблiотечно-бiблiографiчнi та повнотекстовi бази даних у бази знань. Для органiзацiї такої семантичної навігацiї у бiблiотечних iнформацiйних ресурсах необхідно використовувати вiдповiднi структури представлення лiнгвiстичних даних в електронному виглядi. Як вiдомо, найбільш просту i зрозумiлу для користувача реалiзацiю представлення знань забезпечує семантична мережа [7, 19].

За означенням, семантична мережа (СМ) є графом, вершини якого вiдповiдають об'єктам (поняттям), а дуги – зв'язкам мiж цими об'єктами (поняттями). Загалом СМ має забезпечувати [5]: зберiгання вiдомостей про об'єкти (поняття) та про зв'язки мiж ними; пошук об'єктiв (понять) за заданими характеристиками (за зв'язками з iншими об'єктами, поняттями); поповнення i коригування знань системи в процесi навчання (введення нових об'єктiв чи понять, встановлення нових зв'язкiв, вилучення з системи об'єктiв, понять, зв'язкiв); умови здiйснення процедур узагальнення i конкретизацiї знань, а саме, СМ має надавати вiдповiдь на поставлене запитання, знаходити фрагменти СМ (пiдмережi), що вiдповiдають поставленому завданню. Саме за принципом семантичної мережi сьогоднi реалiзуються структури для представлення лiнгвiстичних даних в авторитетних файлах автоматизованих бiблiотечних iнформацiйно-пошукових систем.

В iдеалi класифiкацiйна схема має забезпечувати оптимальний перехiд вiд одного поняття до iншого, вiд рубрики до супровiдної iнформацiї, вiд неї до сумiжних рубрик. Тобто, правильно i послiдовно розроблена класифiкацiйна схема представляє такий рiвень узагальнення знання, який, з одного боку, пiдпорядкований «логiцi свiту», а з другого – логiцi

пошуку [15]. Вiдповiдно, структура представлення класифiкацiйних даних (формальна модель класифiкацiї) має задовольняти такi основнi вимоги: бути зручною для представлення даних у комп'ютерi; адекватно вiдображати будову термiносистеми предметної галузi; створювати умови для повноцiнної та своєчасної актуалiзацiї класифiкацiйних даних; забезпечувати гнучкi сценарiї пошуку та зручний пошуковий користувачський iнтерфейс.

Для вирiшення окресленого кола завдань фахiвцi розробили формальну модель класифiкацiї у виглядi графа, вершинами якого є фрейми, що описують окреми рубрики класифiкатора, об'єднанi в iєрархiчну семантичну мережу [3]. Семантична мережа є статичним компонентом моделi, що репрезентує концепцiю знань, прийнятну в iнформацiйнiй системi, а фрейми рубрик представляють динамiчний термiнологiчний компонент моделi, котрий може ситуативно поповнюватись новими термiнами та термiнологiчними зв'язками. Для динамiчної актуалiзацiї класифiкацiйної схеми фрейми рубрик розширено елементами тезауруса, що враховують нечiткiсть лiнгвiстичних об'єктiв (умовнi синонiми, термiнi-еквiваленти, синонiми iншою мовою) [4]. Така органiзацiя класифiкацiйних даних дає змогу, з одного боку, реалiзувати основну перевагу класифiкацiйної схеми – систематичний тематичний пошук, а з другого – здiйснювати актуальний предметний доступ до iнформацiйних ресурсiв (у т. ч. багатомовний).

Загалом авторитетний запис у базi даних є фреймом-контейнером, де зосереджена вся iнформацiя про те, за якими пошуковими запитаннями можна знайти потрібне поняття. Так, поняття «Нейрокомп'ютери» в Рубрикаторi НБУВ оточене цiлою множиною термiнiв з позначенням взаємовiдносин мiж ними:

Комп'ютери i програмування

↑

Нейрокомп'ютери

нейроннi ЕОМ ; нейроннi комп'ютери

(eng) neural computer

(eng) neurocomputer

✓нейроiнформатика

✓нейрокомп'ютинг

↓

Нейропроцесори

→

Нейроннi мережi

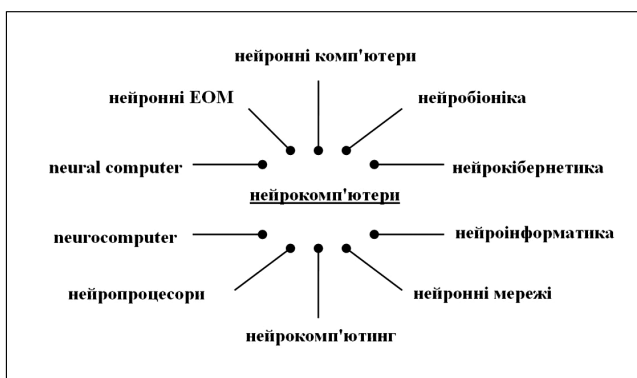
Нейробiонiка

Нейрокiбернетика

Умовнi позначення: ↑ – надрубрика; ↓ – пiдрубрика;

→ – зв'язанi заголовки (асоцiацiї); ✓ – додатковi заголовки.

Результатом упорядкування тематичної інформації є можливість знайти задане поняття за будь-яким терміном із сформованої множини:



Запропоновану модель класифікації було реалізовано в НБУВ на платформі САБ «ІРБІС-64» у вигляді спеціалізованого авторитетного файлу класифікації. За основу представлення інформації в авторитетному файлі взято скорочений варіант полів та підполів «UNIMARC Classification Format» [11], що реалізує структуру фреймів класифікаційних рубрик. Навігаційні засоби системи створюють передумови для укладання міжпредметних зв'язків на основі встановлення асоціативних відношень між поняттями. Розроблена модель та впроваджені технологічні рішення лягли в основу підготовки друкованого та електронного варіанта підрозділів Рубрикатора НБУВ «381 Кібернетика», «397 Інформаційна та обчислювальна техніка» [3, 8], що значно розширило пошукові можливості електронного каталогу НБУВ у галузі комп'ютерних наук та інформатики.

Впровадження електронного варіанта Рубрикатора НБУВ на «Шляху документа» дало змогу реалізувати значні переваги класифікаційної схеми для представлення знань у бібліотечному фонді. Можливості системи забезпечують механізми, необхідні для професійної підтримки рубрикатора бібліотеки в актуальному стані (кумуляція нових термінів, фіксування методичних рішень, введення нових рубрик, встановлення синонімічних та асоціативних відношень). Поповнення рубрикатора у відділі систематизації налаштовано таким чином, що внесені доповнення і зміни, редакційні правки автоматично відображаються у пошуковому інтерфейсі електронного каталогу.

Особливі переваги для користувачів надають навігаційні та лінгвістичні засоби онлайнного електронного каталогу. Веб-інтерфейс рубрикатора дає змогу шукати рубрики за ключовими словами, пересуватися вгору і вниз тематичним деревом рубрикатора, переходити до суміжних розділів знання. Однак, як з'ясувалося, що лише незначна кількість користувачів

(1–2 %) надають перевагу пошуку за рубрикатором. Набагато популярнішим виявився інтерфейс бібліографічних описів, у якому замість класифікаційних індексів виводяться тематичні рубрики у вигляді гіпертекстових посилань, що дають змогу за одним натисканням отримати відповідну тематичну добірку. Цей інтерфейс використовує кожен 3–4 користувач, який надсилає запит до електронного каталогу.

Завдяки лінгвістичним перевагам тематичний пошуковий інтерфейс дає змогу безкоштовно користуватися сервісом Google Translate, який пропонує швидкий автоматичний переклад веб-сторінок багатьма мовами світу. Користувач, вводячи свій запит англійською або російською мовою, може знайти лише частину інформації мовою запиту. Так, за запитом російською мовою «Библия» або англійською «Bible» у полі назва користувач отримує також посилання на тематичну рубрику «Тексти Біблії» (рос. «Тексты Библии», англ. «Bible Texts»), а натиснувши це посилання на знайомій йому мові він отримує добірку всіх біблійних текстів всіма мовами, що зберігаються у фонді НБУВ (українською, російською, англійською, німецькою, польською, французькою тощо). Таку сервісну можливість надає пошук за класифікаційним індексом, бо саме він зберігається у записах і є лінгвістично незалежним.

Аналогічно реалізовано пошук за географічними рубриками відповідно до таблиць територіальних типових поділів (ТПП) Рубрикатора НБУВ [9], електронний варіант яких підключено для здійснення пошуку в онлайнному каталозі та у спеціалізованих базах даних сектора картографії. Наприклад, для того, щоб знайти документи стосовно Австрії, користувачам Австрійської бібліотеки достатньо натиснути рубрику «Австрія» (в перекладі німецькою – «Österreich») і у автоматичному режимі отримати добірку необхідної літератури всіма мовами із фонду НБУВ, надалі є можливість уточнити запит – шукати у знайденому.

Таблиці територіальних типових поділів становлять також основу систематизації картографічного фонду. Для полегшення роботи та інтеграції баз даних картографічного фонду до загальної пошукової системи бібліотеки автоматизоване робоче місце бібліографа було забезпечене багатомовним електронним варіантом таблиць ТПП (українська, російська, англійська мови). На основі конвертованих таблиць організували авторитетний файл «Алфавітно-предметний покажчик ТПП». В електронному алфавітно-предметному покажчику забезпечено можливість багатомовного пошуку та розширений формат перегляду авторитетних записів. Авторитетний файл підключено до робо-

чого аркуша каталогiзатора як довідковий засiб та засiб введення класифiкацiйних iндексiв, вiн також доступний каталогiзаторам для оперативного редагування, поповнення i прийняття методичних рiшень.

Тематичний рубрикатор став зручним iнструментом для створення комплексних iнформацiйних бiблiотечних сервiсiв з єдиною концепцiєю представлення знань, де за натисканням гiперпосилання тематичного роздiлу (категорiї) можна отримати тематичну добiрку iнформацiйних джерел одразу iз всiх баз даних (електронного каталогу, електронної бiблiотеки, iнтернет-навігатора, бiблiографiчної довідки). Всi iнформацiйні продукти представленi в єдиному вiкні створюються багатьма структурними пiдроздiлами бiблiотеки. Для тематичного упорядкування потоку надходжень до фонду електронних iнформацiйних ресурсiв (повнотекстові версiї видань) було організовано їх iндексацiю професiйними систематизаторами, аналогiчно «Шляху документа НБУВ» для паперових документiв. Класифiкацiйними iндексами також надiляються каталогiзованi iнформацiйні ресурси iнтернету. iнтеграцiя комплексного iнформацiйного продукту вiдбувається завдяки використанню єдиного стандарту упорядкування знань – тематичного рубрикатора бiблiотеки. Такий комплексний iнформацiйний сервiс на платформi iнформацiйних продуктiв IPBIS пройшов успішне тестування на порталі НБУВ (irbis-nbuv.gov.ua) i активно використовується користувачами (наприклад упродовж 2011 р. у середньому 400 користувачiв щодоби).

Використання iнструментарiю тематичної фiльтрацiї (заданої добiрки класифiкацiйних iндексiв певної тематики) дало змогу організувати спецiалізовані бiблiотеки, де на кожний запит користувача накладається тематичний фiльтр, який обмежує вiдповiдi на запит заданою предметною галуззю. За такими принципами працюють онлайнові спецiалізовані бiблiотеки «Юридичні iнформацiйні ресурси» (http://irbis-nbuv.gov.ua/info_law.html) та «Фiлософська бiблiотека» (http://irbis-nbuv.gov.ua/info_ph.html), пошуковий iнтерфейс яких супроводжує профiльна частина тематичного рубрикатора (для юридичних iнформацiйних ресурсiв – роздiл X «Державо i право. Юридичні науки»).

Подальшим розвитком спецiалізованих бiблiотек може бути iнтеграцiя у комплексний пошуковий iнтерфейс iнших електронних iнформацiйних ресурсiв НБУВ. Використання рубрикатора для систематизацiї цифрових копiї видань НБУВ в електронному фонді дасть змогу гнучко упорядковувати тематичні електронні колекцiї вiдповiдно до вимог визначеної у бiблiотецi концепцiї представлення знань.

Список використаних джерел

1. Волохiн О. М. Каталогiзацiя цифрових ресурсiв iнтернет : Дублiнське ядро метаданих : посiб. – Кiровоград, 2003. – 70 с.
2. Захаров В. П.; Масевич А. Ц. Актуальные проблемы лингвистического обеспечения автоматизированных систем крупных библиотек России // Информационное обеспечение науки : новые технологии. – М. : Б-ка по естест. наукам, 1997. – С. 143–150.
3. Лобузiна К. Бiблiотечні класифiкацiї в сучасних iнформацiйних сервiсах / К. В. Лобузiна; НАН України, Нац. б-ка України ім. В. І. Вернадського. – К., 2010. – 132 с.
4. Лобузiна К. Класифiкацiйні схеми як засiб представлення знань в електронних бiблiотеках / К. В. Лобузiна // Реєстрацiя, зберiгання i обробка даних. – 2006. – Т. 8. – № 4. – С. 75–86.
5. Любарский Ю. Я. Интеллектуальные информационные системы. – М. : Наука, 1990. – 232 с.
6. Митчелл Д. Десятичная классификация Дьюи: средство организации знаний для следующего века // Библиотеки и ассоциации в меняющемся мире: V Междунар. конф. «Крым-95» : материалы конф. – 1998. – Т. 2. – С. 169–177.
7. Организация знаний: лингвистические аспекты : прогр. и тез. докл. конф. [ИСКО], Москва, 16–18 мая 1995 г. / Междунар. о-во по орг. знаний, Рос. секция, Рос. гос. б-ка. – М. : Рос. секция ИСКО, 1995. – 31 с.
8. Рубрикатор Нацiональної бiблiотеки України iменi В. І. Вернадського. Роздiл «3 Енергетика. Радіоелектронiка». Пiдроздiли «381 Кiбернетика», «397 iнформацiйна та обчислювальна технiка» / Нац. акад. наук України, Нац. б-ка України ім. В. І. Вернадського – К., 2011. – 249 с.
9. Рубрикатор Нацiональної бiблiотеки України ім. В. І. Вернадського. Таблиці територiальних типових подiлiв / І. І. Багрій [та iн.] ; НАН України, Нац. б-ка України ім. В. І. Вернадського. – К., 2004. – 768 с.
10. Berners-Lee T., Hendler J., Lassila O. The Semantic Web : A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities [Electronic resource] // Scientific American. – 2001. – URL : <http://www.sciam.com>. – Last revised : May 17, 2001.
11. Concise UNIMARC Classification Format (20001031) // International Federation of Library Associations and Institutions (IFLA) [Електронний ресурс] / Way of access : URL : <http://www.ifla.org/VI/3/p1996-1/concise.htm>. – Last revised: November 24, 2000.
12. DDC database // Classification research for knowledge representation and organization : Proceedings of the 5th international study conference on classification research, Toronto, Canada, 24–28 June 1991 / edited by N. J. Williamson and M. Hudon. – New York : Elsevier, 1991. – P. 29–37.
13. Dublin Core Metadata Initiative [Electronic resource]. – URL : <http://dublincore.org>.
14. Hodge G. Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. – Washington : The Digital Library Federation Council on Library and Information Resources, 2000. – 43 p.

Питання кирилично-латиничної транслітерації у контексті систематизації бібліографічних даних

15. *Lancaster F. W.* Vocabulary control for information retrieval. – Washington, D. C. : Inf. Resources Press, 1972. – 233 p.

16. Master Reference File // UDC Consortium [Electronic resource]. – Way of access : URL : <http://www.udcc.org/mrf.htm>.

17. The Role of classification schemes in Internet resource description and discovery // Work Package 3 of Telematics for Research project DESIRE (RE 1004) [Electronic resource]. – URL : <http://www.ukoln.ac.uk/metadata/esire/classification/>. – Last revised : Jan 28, 1999.

18. USMARC format for classification data : including guidelines for content designation. – Washington, D. C. : Library of Congress Cataloging Distribution Service, 1991.

19. *Wang J. A.* Knowledge Network Constructed by Integrating Classification, Thesaurus, and Metadata in Digital Library // Intl. Inform. & Libr. Rev. – 2003. – № 35. – P. 383–397.

20. The World Wide Web Consortium [Electronic resource]. – URL : <http://www.w3.org/>.