

## **РОЗШИРЕННЯ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНОГО ЗАБЕЗПЕЧЕННЯ УНІВЕРСИТЕТСЬКИХ БІБЛІОТЕК ШЛЯХОМ ВВЕДЕННЯ БАЗ ЗНАНЬ**

Університетські та публічні бібліотеки в Україні, Росії та інших країнах СНД використовують ряд систем для обслуговування користувачів в автоматизованому режимі. У даному дослідженні ми проаналізували наступні системи: “ALEPH”, “MARK-SQL”, ІРБІС, “МегаПро”, LIBER MEDIA, “УФД/Бібліотека” [1,2]. Виконаний аналіз показує, що найбільш поширеними є системи ІРБІС і «УФД/Бібліотека».

Дані системи є типовими інтегрованими рішеннями в області автоматизації бібліотечних технологій і призначені для використання в бібліотеках будь-якого типу й профілю. Такі системи цілком відповідають міжнародним вимогам до подібного роду систем, і підтримують усі вітчизняні бібліографічні стандарти й формати.

У даних системах реалізовані всі типові бібліотечні інформаційні задачі, включаючи технології комплектування, систематизації, каталогізації, читацького пошуку, книговидачі й адміністрування.

Системи мають ряд прикладних функцій: комплектування літератури; облік бібліотечного фонду; створення і ведення електронного каталогу; систематизація та предметизація документів; технічна обробка нових видань; довідково-інформаційне обслуговування; обслуговування читачів; ведення статистичної звітності; аналіз книгозабезпеченості; робота з повнотекстовими та мультимедійними ресурсами; застосування технології штрих-кодів; доступ до ресурсів бібліотеки через Інтернет.

Існуючі системи забезпечують комплексну автоматизацію інформаційних бібліотечних задач, ефективне управління інформаційними ресурсами і організацію доступу до них. Функціональні можливості систем дозволяють підтримку форматів MARC21, RUSMARC, UNIMARC; підтримку UNICODE на рівні зберігання та представлення даних; підтримку різних видів лінгвістичного забезпечення; пошук по різних елементах бібліографічних записів; навігацію по функціях та інформаційних ресурсах систем; підтримку довільної кількості бібліографічних баз даних; обслуговування читачів, відповідність вимогам до захисту персональних даних; функціонування в локальних і глобальних мережах; інтеграція в корпоративні системи на основі протоколів Z39.50, HTTP, OAI – PMH.

Дані системи є інтегрованими, тобто видом бібліотечно-програмного забезпечення яке має гнучку адаптацію до потреб конкретної бібліотеки і забезпечує комплексну автоматизацію всіх процесів бібліотечної технології. Інструментальні засоби систем дають змогу підтримувати графічні зображення і повні тексти документів; MARC-сумісні формати; ISO-стандарти; інформаційно-пошукові мови дескрипторного й класифікаційного типів з наявною в них системою посилань; файли авторизації; штрихові коди (бар-коди); протоколи передачі даних для роботи в локальних, корпоративних і глобальних мережах; WWW-сервер, який надає широкий спектр послуг через Інтернет.

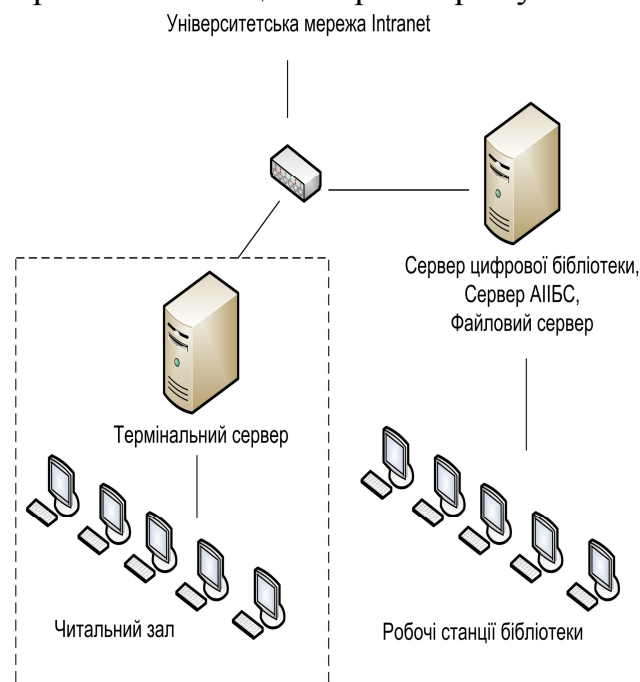
**Метою** пропонованих тез доповіді щодо виконуваного дисертаційного дослідження є формування необхідних формально-логічних основ для побудови інтелектуального рішення в рамках існуючого інформаційно-програмного забезпечення університетських бібліотек України та для забезпечення оптимального підбору навчально-методичного забезпечення процесу вивчення фахових навчальних дисциплін з використанням технології баз знань та методів добування даних.

Розгляд питань пов'язаних з логічною структурою програмних комплексів, в нашому випадку це цифрова бібліотека, доцільно розглядати в межах інтеграції системи в існуючий машино-програмний комплекс. Пропоноване рішення побудови цифрової бібліотеки базується на існуючому технічному забезпеченні науково-технічної бібліотеки і вимагає мінімальних матеріальних витрат на її розробку та впровадження [3]. Принципова схема мережі бібліотеки представлена на рис.1.

Таке рішення забезпечує централізовану підтримку таких функцій:

- адміністрування;
- наповнення бібліотеки;
- надання послуг (сервісів);
- супровід програмного забезпечення.

Таким чином, наповнення цифрової бібліотеки контентом може здійснюватися користувачами як в межах самої бібліотеки так і віддаленими користувачами на своїх робочих станціях через мережу Internet/Intranet.



**Рис. 1 Принципова схема мережі бібліотеки**

Структура цифрової бібліотеки повинна володіти наступними властивостями:

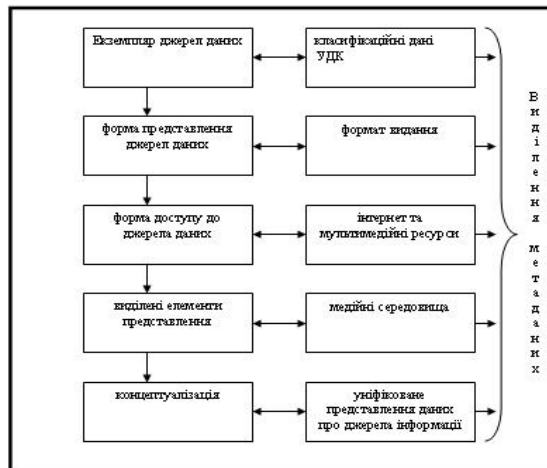
- централізоване ведення репозиторіїв та надання послуг;
- надання можливостей віддаленого наповнення цифрової бібліотеки інформаційними ресурсами;
- адміністрування прав доступу до системи;

- створення електронних архівів інформаційних ресурсів різних видів;
- надання функцій інтелектуального пошуку (повнотекстового та розширеного з використанням метаданих) і перегляду (в тому числі медіа-ресурсів);
- використання міжнародних стандартів (розширене Дублінське ядро) і протоколу OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting) при збереженні описів публікацій (метаданих);
- надання можливості для інтеграції з іншими каталогами, реєстрами та репозиторіями;
- індексація Web-сервісу пошуковими системами.

Проектована система підтримки прийняття рішень при виборі навчальних та навчально-методичних матеріалів в бібліотеці університету базується на методології експертних систем для того щоб мати достатню функціональність для управління процесами обробки даних в контексті вирішення задачі класифікації ресурсів університетської бібліотеки. Тому, основне завдання такої системи є побудова рекомендацій щодо методу добування домінантних особливостей того чи іншого навчального джерела згідно множини оцінок заданих експертами (викладачами навчальних курсів), а також задання класифікатора для заданого набору даних відповідно до наявної множини правил, що описують певну навчальну проблему, яка вирішується при вивченні відповідної теми з навчального модуля.

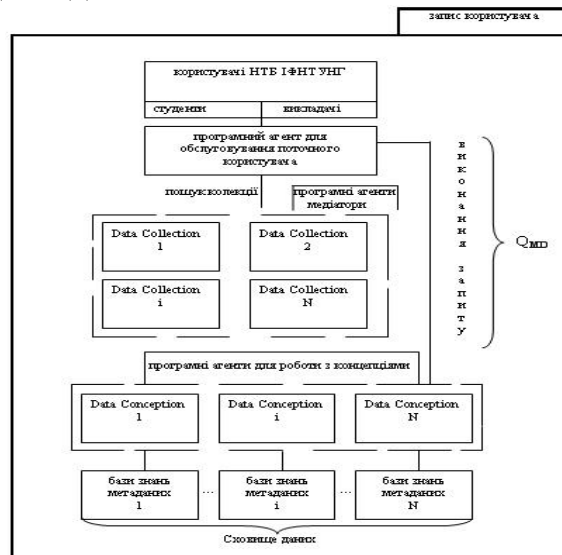
Одним із ключових питань з точки зору математичного моделювання процесу побудови інформаційних інтелектуальних систем для бібліотечної справи є спосіб представлення знань, на основі якого система повинна приймати рішення в певній ситуації. Таким чином представлення знань повинно бути задано способом, який дозволяє перехід до представлення фрагментів інформації про бібліотечний ресурс в термінах структур баз знань (БЗ), зокрема баз знань метаданих як засобу підтримки логічного виведення на множині метаданих. Розглядатимемо таку базу знань з точки зору фактів і процесів, що призводять до їх зміни, тобто з погляду семантики і синтаксису такого представлення. Під синтаксисом будемо розуміти набір правил для поєднання символів в логічно коректні вирази, а під семантикою – спосіб інтерпретації виразів, що одержуються в результаті конкретних реалізацій синтаксичних правил [4].

Ми використовуємо метадані, що базуються на формальних онтологічних принципах підтримки належного обчислювального обґрунтування. Формальні онтологічні принципи використовують логіку для визначення зв'язків з іншими концепціями. Тому множину інформаційних ресурсів бібліотеки ми розглядаємо виходячи з множини онтологічних концепцій. Це дає можливість досліджувати зв'язки між властивостями, що визначені в онтології, а також зв'язків з бібліотечними входженнями. Так, наприклад, можна визначити, коли одна послуга є специфікацією іншої, навіть якщо це співвідношення не представлено у визначенні. Тому заслуговує уваги виділення та представлення у вигляді БЗ множини метаданих цифрової бібліотеки.



**Рис. 2 Робоча ієрархія і зв'язані атрибути**

Формальна онтологія є відповідною технікою для моделювання комплексних доменів. Визначені концепції можуть утворювати мережі відносин без прив'язки до моделюючого інформаційного дерева. Концепції можуть мати багато описових вимірів (атрибутів), можуть бути також частково описані на будь-якому рівні концептуалізації (з будь-якою комбінацією вимірів) і можуть розглядатися з багатьох перспектив (оцінюватися різними послідовностями атрибутів). Вибір певного входження в онтології дозволяє підтримувати доступ з будь-якої перспективи на будь-якому рівні концептуалізації. Для порівняння, декларативні формалізми з меншою виразністю, такі як бази даних, створюють основу для побудови комбінацій та класифікацій вимірів відповідної бази знань.



**Рис. 3 Використання програмних агентів**

Розглядатимемо БЗ метаданих онтології, як набір інформаційних сутностей атомарних предикатів з деякого скінченного інформаційного простору  $O$ . Всі зміни, що відбуватимуться в БЗ, будемо розглядати, як наслідок запитів користувача по бібліотечним ресурсам, що генеруються. Основою самих запитів є набір модифікаційних предикатних правил. Розглядатимемо два типи правил:

$$KB_{MD+}(o) \ll KB_{MD+}(o_1), \dots, KB_{MD+}(o_l),$$

$$KB_{MD-}(p_1), \dots, KB_{MD-}(p_m)$$
(1)

$$KB_{MD-}(o) << KB_{MD+}(o_1), \dots, KB_{MD+}(o_l), \quad (2)$$

$$KB_{MD-}(p_1), \dots, KB_{MD-}(p_m)$$

де  $o, o_i, p_i \in O$ . Основна ідея такого запису правил полягає в тому, що  $KB_{MD+}(o)$  означає, що атомарний предикат  $o$  повинен бути включений в БЗ  $KB_{MD}$ , а  $KB_{MD-}$  означає, що  $o$  – повинен бути виключений з бази знань.

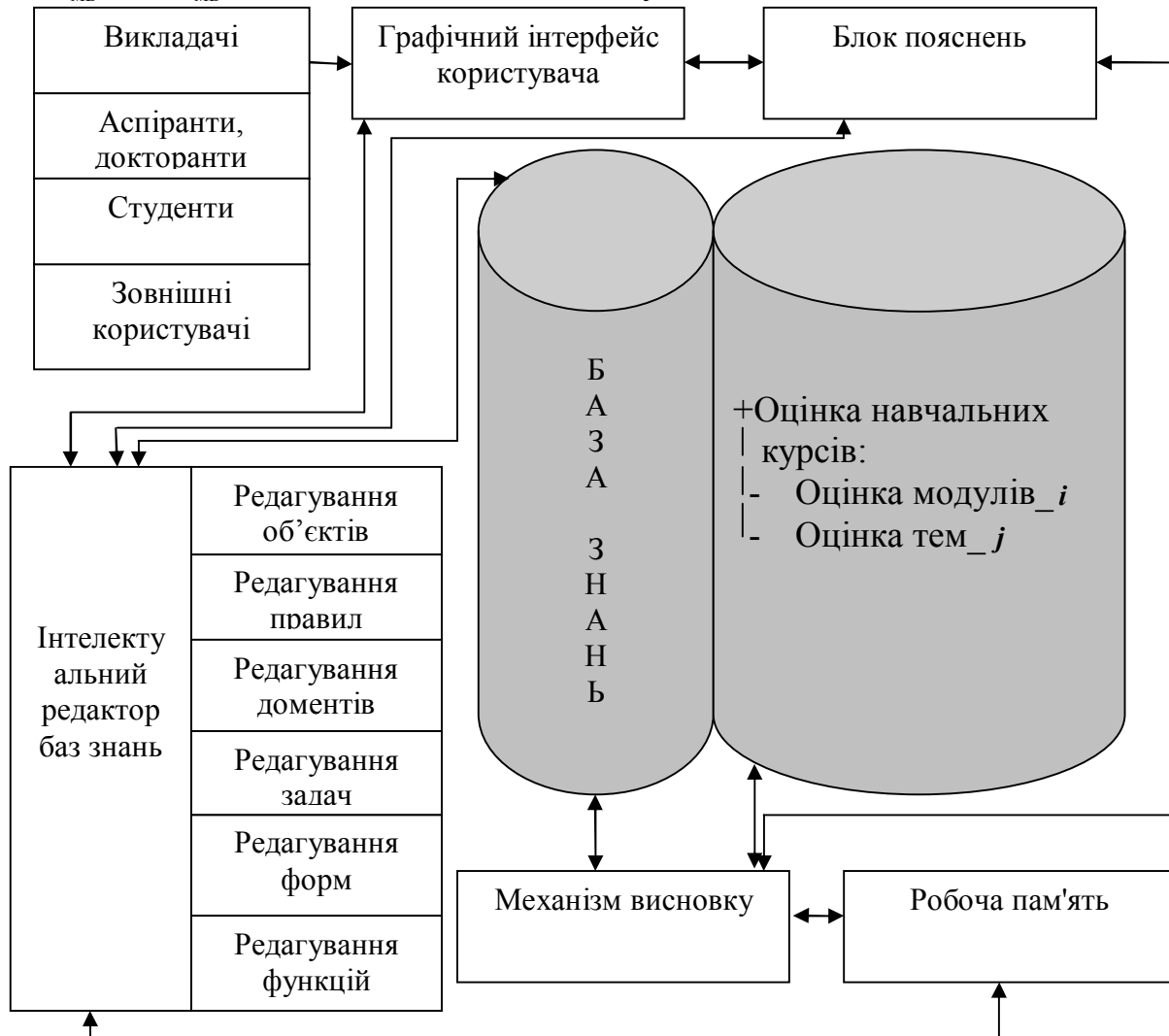


Рис. 4 Структура інтеграції баз знань у інформаційному середовищі бібліотеки Семантично ініціалізації БЗ метаданих  $KB_{MD}^{init}$  позначають об'єкти, концепції позначають множини об'єктів і відношення позначають множини (кортежі) об'єктів [5]:

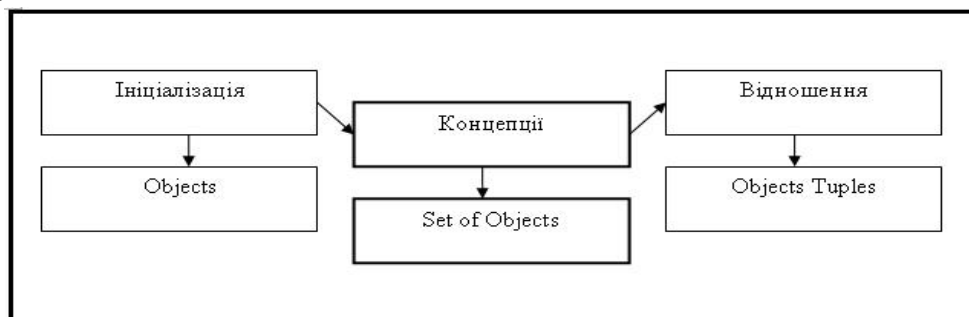


Рис. 5 Семантична ініціалізація баз знань метаданих

**Означення 1.** Будемо вважати, що концепція  $Conception^1$  для  $KB_{MD}$  продукує іншу концепцію  $Conception^2$  якщо кожна ініціалізація структури бази знань метаданих теж є відповідною продукцією.

**Твердження 1.** Модифікація  $Conception^1 \rightarrow Conception^2$  має місце тоді і тільки тоді якщо  $[Conception^2]^{Models} \subseteq [Conception^1]^{Models}$  для всіх логічних моделей бази знань метаданих  $Models_{KB_{MD}}$ , що задають відображення введених символів на універсум об'єктів предметної області, де  $[Conception^2]^{Models}$  і  $[Conception^1]^{Models}$  є розширеннями (тобто множинами об'єктів для цих концепцій).

**Означення 2.** Значення концепцій можна інтрептувати функцією, що задає відображення з множини логічних моделей на множини розширень концепцій  $f_{Conception} : Models_{KB_{MD}} \rightarrow Conception^{Models_{KB_{MD}}}$ .

**Означення 3.** Формальну онтологію  $FO$  будемо розглядати як множину обмежень  $Constr_{set}$ , що накладена на множину можливих моделей  $Models_{KB_{MD}}$ .

**Означення 4.** Нехай  $Q_{MD}$  - модифікаційний запит по базі знань метаданих у виділеній концепції  $SelectedConception$ , а  $KB_{MD1}$  і  $KB_{MD2}$  дві БЗ. Означимо залишок для  $Q_{MD}$  стосовно  $(KB_{MD1}, KB_{MD2})$  як  $Q_{MD}^{KB_{MD2}}|KB_{MD1}$ , який одержується в результаті:

- 1) видалення з  $Q_{MD}$  кожного модифікаційного правила, body-частина якого не співставляється з  $KB_{MD2}$  (в результаті одержимо  $Q_{MD}^{KB_{MD2}}$ ) згідно введеної множини  $Constr_{set}$ .
- 2) видалення кожного модифікаційного літералу з body-частин правил в  $Q_{MD}^{KB_{MD2}}$ , що співставляється з  $KB_{MD1}$  в множині  $Constr_{set}$ .

**Означення 5.** Нехай  $Q_{MD}$ -модифікаційний запит по базі знань метаданих у виділеній концепції  $SelectedConception$  і  $KB_{MD1}$ ,  $KB_{MD2}$  — дві БЗ. Тоді справедливо:

$$\lambda_{nm}(Q_{MD}^{KB_{MD1}, KB_{MD2}}) \text{ є когерентним і } KB_{MD2} = KB_{MD1} \circ \lambda_{nm}(Q_{MD}^{KB_{MD1}, KB_{MD2}})|_{Constr_{set}},$$

$$\lambda_{nm}(Q_{MD}^{KB_{MD2}}|KB_{MD1}) \text{ є когерентним і } KB_{MD2} = KB_{MD1} \circ \lambda_{nm}(Q_{MD}^{KB_{MD2}}|KB_{MD1})|_{Constr_{set}}.$$

**Означення 6.** Нехай  $Q_{MD}$ -модифікаційний запит по базі знань метаданих у виділеній концепції  $SelectedConception$  і  $KB_{MD2}$  є  $Q_{MD}$ -модифікацією для  $KB_{MD1}$  в заданій системі обмежень  $Constr_{set}$ . Тоді

$$\lambda_{nm}(Q_{MD}^{KB_{MD1}, KB_{MD2}})|_{Constr_{set}} = \lambda_{nm}(Q_{MD}^{KB_{MD2}}|KB_{MD1})|_{Constr_{set}} = head(Q_{MD}^{KB_{MD2}}).$$

**Означення 7.** Виходячи із вищесказаного, справедливим буде наступне :

1) БЗ  $KB_{MD2} \in Q_{MD}$ -модифікацією БЗ  $KB_{MD1}$  в заданій системі обмежень  $Constr_{set}$

$$2) \lambda_{nm}(Q_{MD} \cup \{o \mid SelectedConception \ll o \in O : (KB_{MD1}, KB_{MD2})\}) \subset KB_{MD2}.$$

$$3) \lambda_{nm}(Q_{MD}^{KB_{MD1}, KB_{MD2}}) \cup O_I(KB_{MD1}, KB_{MD2}) \subset KB_{MD2}.$$

**Означення 8.** Нехай  $Q_{MD}$ -модифікаційний запит по базі знань метаданих у виділеній концепції *SelectedConception* і  $KB_{MD1}$ — БЗ. Якщо БЗ  $KB_{MD2} \in Q_{MD}$ -модифікацією для  $KB_{MD1}$ , тоді  $KB_{MD2}$  є моделлю  $Model_{KB_{MD}}$  для  $Q_{MD}$ .

Введене нами означення модифікаційного запиту задовольняє також принцип мінімальності. Тобто, якщо виконати  $Q_{MD}^{Constr_{set}}$ -модифікацією для  $KB_{MD1}$ , то одержана БЗ  $KB_{MD2}$  відрізнятиметься від вихідної мінімально. Для вимірювання цієї характеристики використаємо принцип симетричної різниці для БЗ. А саме:

$$dist(KB_{MD1}, KB_{MD2}) = (KB_{MD1} \setminus KB_{MD2}) \cup (KB_{MD2} \setminus KB_{MD1})|_{Constr_{set}}$$

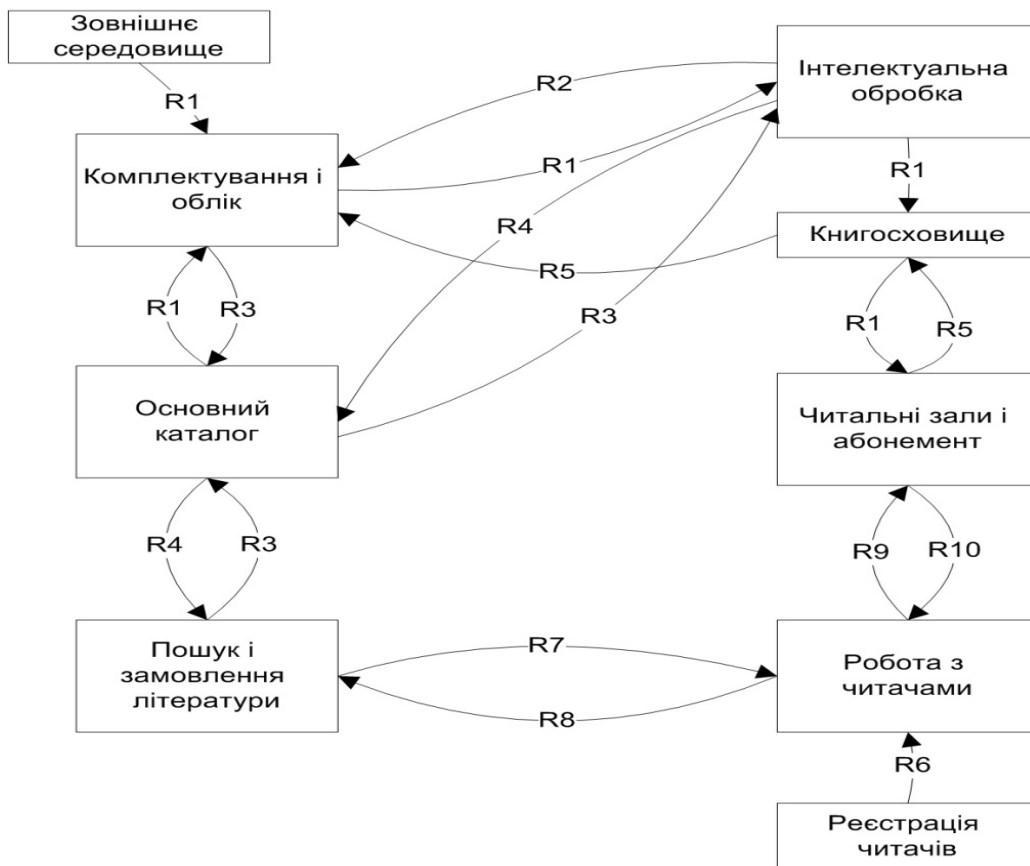
**Означення 9.** Добування бібліотечних даних, документів та інформації будемо розглядати у формі стратегії добування  $RS$  при якій метрична міра оцінки  $EM$  дозволяє визначити ступінь схожості (метричної близькості) між множинами документів добутих у стратегії  $RS$  і множиною релевантних документів  $RD^{set} \rightarrow [KB : RD^{set}]$ , яка внесена у базу знань системи експертом предметної області наприклад, лектором профільної кафедри вузу.

**Означення 10.** Запитом у бібліотечній системі університету будемо вважати певне формулювання інформаційної потреби студента. В літературних джерелах [6-8] найпростішим видом запитів вважаються запити на основі ключових слів, які у своїй природі є інтуїтивними, простими у вираженні, дозволяють швидко ранжування і, як наслідок, система знаходить всі документи, що містять задані ключові слова. Тому запити на основі одного або кількох ключових слів розглядаються як базові. Результатом таких запитів є множина, яка містить хоча б одне з ключових слів, вказаних в запитах. На наступному кроці результуючі документи ранжуються відповідно до ступеня схожості до кінцевого запиту. Основа механізму ранжування результатів полягає в використанні характеристики частоти входження слова у документ та інверсної документної частоти, яка вказує кількість документів, у яких з'являється задане слово. У даному контексті також слід враховувати проблеми багатозначності слів.

Представлена в роботі структура цифрової бібліотеки і висунуті вимоги до функціонування такої системи, дають змогу представити дану систему у вигляді моделі взаємодії підсистем ЦБ за допомогою орієнтованого графа (рис. 6):

$$G = (V, A)$$

Інформаційний простір формують такі процеси (сервіси) та об'єкти (вершини орграфа – множина  $V$ ): комплектування та облік; інтелектуальна обробка; книгосховище; основний каталог; пошук і замовлення літератури; робота з читачами; читальні зали і абонемент; реєстрація читачів; робота з читачами. А також наступні комунікації (ребра орграфа – множина  $A$ ):  $R1$  – надходження літератури;  $R2$  – узгодити ідентифікатори карток каталогу;  $R3$  – знайти літературу;  $R4$  – отримати знайдену літературу;  $R5$  – списати літературу;  $R6$  – реєстрація нового читача;  $R7$  – здійснити замовлення літератури;  $R8$  – читач є боржником;  $R9$  – обслужити читача;  $R10$  – читач повернув літературу.

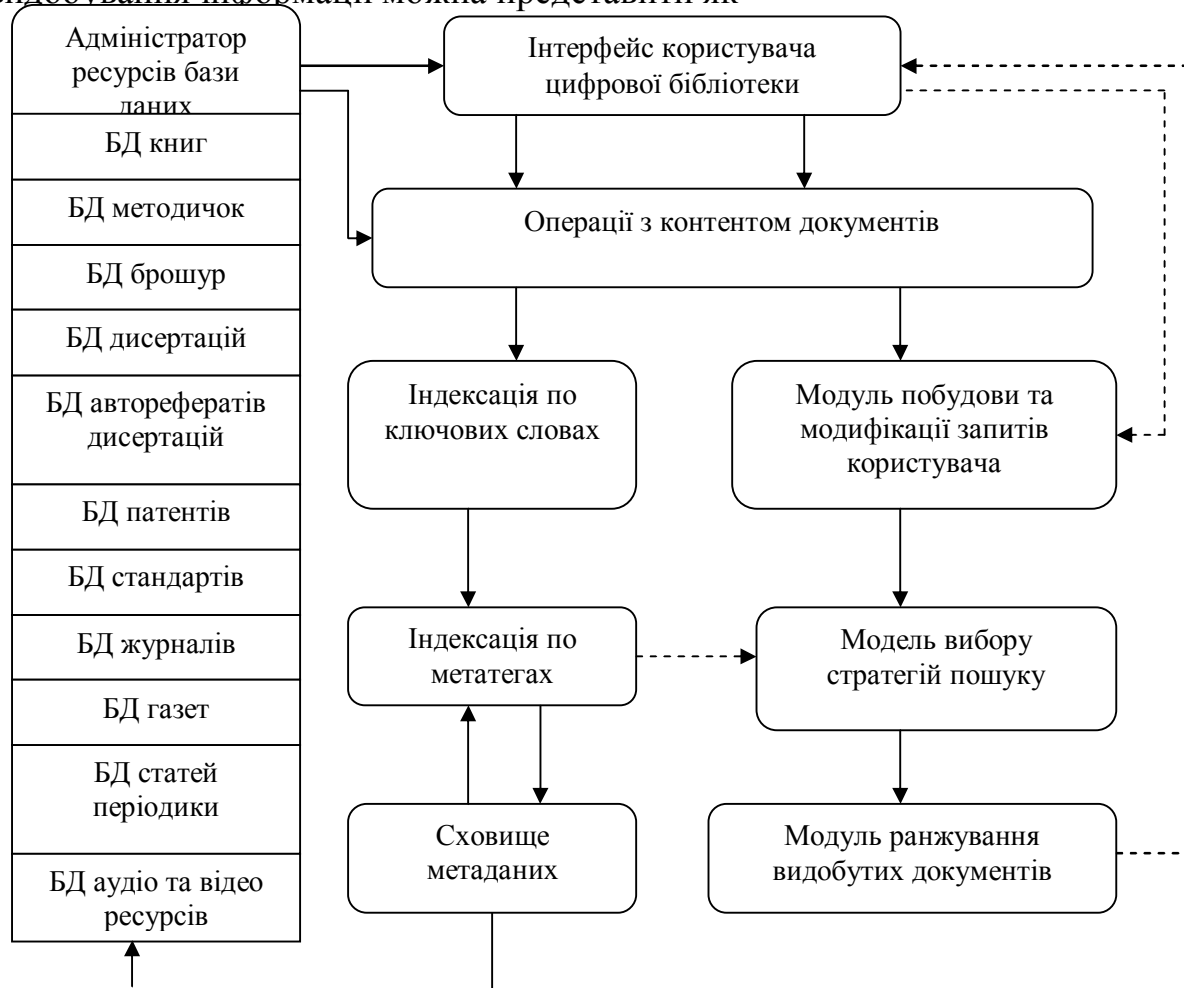


**Рис. 6 Орієнтований граф взаємодії процесів в інформаційному просторі**

Для інтеграції інформаційного простору бібліотеки в єдиний інформаційний простір університету, необхідно забезпечити взаємодію АБС з системами адміністративно-організаційного управління, а саме – системами відділу кадрів, бухгалтерії і деканатів. Це дозволить спростити роботу бібліотекарів з читачами, які є боржниками, спростити процедуру підпису обхідних листів, а також забезпечити автоматичну реєстрацію нових працівників вузу та студентів [9].

Важливу роль в навчальному процесі Івано-Франківського національного технічного університету нафти і газу відіграє науково-технічна бібліотека. Основною задачею розвитку її інформаційних ресурсів на сьогоднішній день є задача побудови інтелектуальної інформаційної системи підтримки цифрової

бібліотеки [10]. Представлений підхід в роботі дозволяє виконувати інтерпретацію даної задачі на рівні метаданих і відповідно структуру процесу видобування інформації можна представити як



**Рис. 7 Структуризація процесу видобування інформації**

Для практичної реалізації даної задачі необхідно детально проаналізувати процес видобування інформації в WEB та в існуючому інформаційному програмному забезпеченні бібліотеки. Важливими умовами даного процесу є: 1) виділення документів для використання; 2) визначення операцій, що будуть виконуватись над текстовим контентом; 3) визначення структури текстового контенту. Таким чином кожна операція над контентом використовуватиме перетворення початкових концептів і генерацію відповідного логічного представлення. Після того як визначено логічне представлення документа система виконує побудову систем індексації. В якості основних ресурсів виступає час і доступний дисковий простір необхідний на визначення текстового контенту БД і відповідно процедура побудови індексів базуватиметься на процедурі побудови запитів для процедури видобування. Враховуючи, що БД документів є індексованими можна ініціалізувати процедуру видобування інформації. Для цього виконуються такі кроки: 1) користувач визначає та специфікує власні інформаційні потреби; 2) використовується парсинг і трансформація визначеної інформаційної потреби на основі операцій визначених для текстового контенту; 3) застосовуються операції перетворення запиту у процесі генерації логічного представлення

інформаційної потреби користувача; 4) виконується обробка запиту з метою отримання видобутих документів. Швидкість обробки запитів визначається вибраною структурою індексації. Перед відправкою користувачу видобуті документи ранжуються відповідно до імовірності їх релевантності інформаційним потребам користувача. Кінцеве сортування виконує користувач будуючи кінцеву множину ранжованих документів. Інтелектуальна система використовує документи вибрані користувачем для зміни сформованого запиту. Кожна послідовна модифікація запиту буде відповідати кращому представленню його інформаційних потреб. В даному дослідженні представлено аналіз підходів до побудови інтелектуальних методів видобування релевантної інформації та документів.



**Рис. 8 Структура застосування процесу добування даних в бібліотеці університету**

Згідно концепції Болонського процесу навчальний курс складається з змістовних модулів  $LM_j$ , які відображають відповідно певні концептуальні напрями навчального курсу. В реальних навчальних курсах, що читаються у національному університеті нафти і газу, кількість змістовних модулів варіюється від одного до кількох. У свою чергу, кожен навчальний модуль складається з сукупності навчальних елементів  $LE_k$ , які по суті є переліком тем (або підтем) навчального модуля. На рівні формального представлення це зобразимо у вигляді  $LC_i = \{LM_j\}$ ,  $LM_j = \{LE_k\}$ ,  $i, j, k \in N$ .

Така структуризація необхідна нам для введення підходу контролю знань студента та формування відповідних проєкцій його зони найближчого розвитку по введених вищих характеристиках. Тобто, у термінах множин ключових слів, що виражають сутність навчального матеріалу на рівні навчального курсу, навчальних елементів та змістовних модулів, отримаємо наступне представлення

$$LC_i \rightarrow KeyWords_{LC_i}^{set}, LM_j \rightarrow KeyWords_{LM_j}^{set}, LE_k \rightarrow KeyWords_{LE_k}^{set}$$

$$PDZ \cong [KeyWords_{LC_i}^{set}] = \\ = \bigcup_j \left\{ KeyWords_{LM_j}^{set} = \bigcup_k [KeyWords_{LE_k}^{set}] \right\}_{j,k \in N}$$

Таким чином, проєктований модуль забезпечуватиме певне попереднє тестування знань студента для визначення рівня його володіння множинами ключових слів заданих рівнів та формування на їх основі проєкцій його зони найближчого розвитку. Відповідно лектор як автор навчального курсу забезпечує ранжування навчально-методичного забезпечення по введених рівнях представлень множинами ключових слів.

### Список літератури

1. <http://library.kr.ua/irbisu.html>.
2. <http://www.matriks-pres.com.ua>.
3. Стисло Т. Структуризація інформаційних ресурсів цифрової бібліотеки університету / Тарас Стисло // Комп'ютерні науки та інформ. технології: матеріали 4-ї Міжнар. наук.-техн. конф. CSIT-2009, Львів, 15-17 жовт. 2009р. – Львів, 2009. – С. 367-370.
4. Стисло Т. Підтримка запитів користувача по бібліотечних ресурсах на основі баз знань метаданих / Тарас Стисло // Комп'ютерні науки та інженерія: матеріали III Міжнар. конф. молодих вчених CSE-2009, Львів, 14 – 16 трав. 2009р. – Львів, 2009. – С. 86-88.
5. Стисло Т.Р. Концептуалізація запитів користувача по базах знань метаданих про бібліотечні ресурси / Т.Р. Стисло, В.І. Шекета, Р.М. Федорак // Вісн. Хмельниц. нац. ун-ту. Техн. науки. – 2009. - №4. - С.132-138.
6. Korfhage R. Information Storage and Retrieval / Robert Korfhage. - New York: John Wiley & Sons, Inc., 1997. – 510 p.
7. Shaw Jr M., Burgin R., Howell P. Performance standards and evaluations in IR test collections: Vector-space and other retrieval models. *Information Processing & Management*, (1997), pp. 15-36.
8. Tague – Sutcliffe J. Measuring the informativeness of a retrieval process, in *Proc. of the 15<sup>th</sup> Annual Int. ACM SIGIR Conference on R&D in Information Retrieval*, Copenhagen, Denmark, (1992), pp. 23-36.
9. Стисло Т.Р. Структуризація інформаційного простору цифрової бібліотеки з допомогою орієнтованих графів / Стисло Тарас // Теоретичні та практичні аспекти побудови програмних систем (ТАAPSD'2009) : тези доп. шостої міжнар. конф. (8-10 груд. 2009). – К., 2009. – С. 108-109.
10. Стисло Т.Р. Структуризація процесу видобування інформації в цифровій бібліотеці університету / Т.Р. Стисло, О.Р. Стисло // Математичне та програмне забезпечення інтелектуальних систем (MPZIS-2010) : тези доп. VIII міжнар. наук.-практ. конф. – Дніпропетровськ : Нац. ун-т ім. Олеся Гончара, 2010. – С. 218-219.