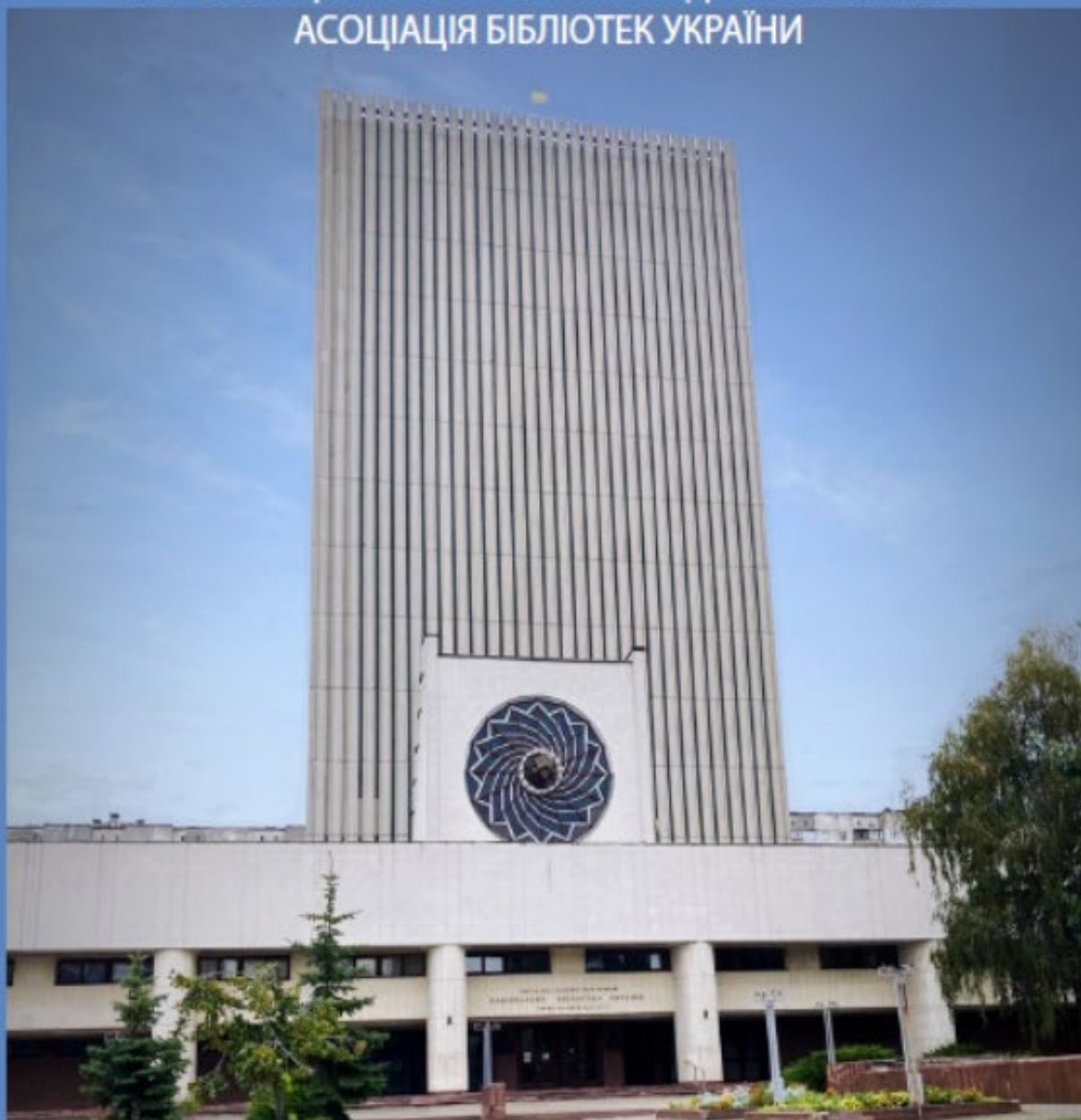


НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
НАЦІОНАЛЬНА БІБЛІОТЕКА УКРАЇНИ
імені В. І. ВЕРНАДСЬКОГО
ІНФОРМАЦІЙНО-БІБЛІОТЕЧНА РАДА НАН УКРАЇНИ
АСОЦІАЦІЯ БІБЛІОТЕК УКРАЇНИ



БІБЛІОТЕКА. НАУКА. КОМУНІКАЦІЯ.
ПРІОРИТЕТИ СЬОГОДЕННЯ
ТА ПЕРСПЕКТИВИ

Матеріали міжнародної наукової конференції
(Київ, 1–3 жовтня 2025 р.)

Том 1

Ланде Дмитро Володимирович,
ORCID <https://orcid.org/0000-0003-3945-1178>,
доктор технічних наук, професор,
завідувач кафедри,
кафедра інформаційної безпеки,
Київський політехнічний інститут імені Ігоря Сікорського,
Київ, Україна
e-mail: dwl@gmail.com

Крючин Андрій Андрійович,
ORCID <https://orcid.org/0000-0002-5063-4146>,
член-кореспондент Національної академії наук України,
доктор технічних наук, професор,
заступник директора,
Інститут проблем реєстрації інформації Національної академії наук України,
Київ, Україна
e-mail: kryuchyn@gmail.com

Застосування великої мовної моделі у технології традиційного пошуку в реферативних базах даних

Представлено інтеграцію великої мовної моделі (LLM) у пошукову систему «Бібліотека Наукометрії», яка об'єднує реферати українських наукових публікацій з РБД «Україніка наукова» та препринти з ArXiv. Основним пошуковим двигуном є Manticore Search, до якого додано семантичне ядро на основі LLama-4 через API. Запропоновано математичне обґрунтування семантичного індексування текстів, а також описано режими роботи, що розширюють можливості аналізу: формування дайджестів, переклад, інтерактивна мережа понять. Показано ефективність системи у пошуку за темами, виявленні трендів, побудові авторських зв'язків. Точність пошуку зросла на 15%, час аналізу скоротився на 30–40%.

Ключові слова: Велика мовна модель, LLM, Manticore Search, семантичний пошук, реферативна база даних, дайджести, інтерактивна семантична мережа, ArXiv, Україніка наукова, наукометрія.

Сучасні науково-технічні бібліотеки зіштовхуються з проблемою обробки все зростаючого потоку інформації. Традиційні методи пошуку, що спираються на ключові слова та частотні моделі тексту, стають недостатніми для ефективного аналізу складних наукових документів. База даних «Бібліотека Наукометрії» [<https://bigsearch.space>] охоплює дві основні колекції: реферати українських наукових публікацій із РБД «Україніка наукова» (1998–2024 рр., 769 тисяч записів) та препринти з ArXiv (2007–2025 рр., 2247 тисяч документів). Для поліпшення пошуку та аналізу інформації в цих масивних корпусах застосовано нову архітектуру, яка поєднує традиційний двигун Manticore Search з сучасною великою мовною моделлю (Large Language Model, LLM), а саме – LLama-4 через API. Такий підхід забезпечує не лише точне знаходження відповідей, а й глибокий семантичний аналіз змісту.

На практиці існуюча система показала ефективність у таких задачах: пошук за темою тепер дозволяє знаходити документи, пов'язані з конкретною проблематикою, навіть якщо вони не містять точно співпадаючих слів, завдяки використанню векторного простору і косинусної подібності. Система також успішно визначає динаміку наукових досліджень у часі, виявляючи тренди та зміни в напрямках публікацій. Особливо корисним виявився механізм побудови зв'язків між авторами,

ключовими словами та темами, що дозволило виявити наукові мережі співпраці та напрямки досліджень.

Інтеграція LLM у пошукову інфраструктуру реферативних баз даних є трендом, який активно розвивається. Проте більшість існуючих систем використовують LLM лише для генерації відповідей, не впроваджуючи глибокого аналізу структури знань. У нашому випадку вдалося реалізувати комплексний підхід, де LLM не замінює традиційні методи, а доповнює їх, забезпечуючи семантичне розширення. Це відкриває нові можливості для наукового пошуку, прогнозування трендів, аналізу авторських мереж та формування освітніх маршрутів.

Протягом останніх років спостерігається значний інтерес до застосування великих мовних моделей у науково-інформаційних системах. Розвиток технологій обробки природної мови, зокрема архітектур на основі трансформерів, відкрив нові можливості для аналізу текстових корпусів, що вимірюються мільйонами документів. Серед основних праць у цій галузі виділяються дослідження, присвячені семантичному пошуку, автоматичному реферуванню, побудові тематичних моделей і мереж знань.

Одним із перших піонерів у сфері застосування LLM у науковому контексті став проєкт Semantic Scholar, який використовує навчальні моделі для формування графів понять і рекомендацій наукових публікацій [1-2]. Важливим етапом стало впровадження моделей BERT і її наукової версії SciBERT [3], які показали високу ефективність у завданнях класифікації, вилучення фактів і векторного представлення наукового тексту. Пізніше появились більш потужні архітектури, такі як GPT-4 і LLaMa-4, які не лише забезпечили кращу якість обробки, але й відкрили шлях до генерації узагальнень, перекладів та інтерактивного аналізу.

У роботі [4] було показано, що навчання мовних моделей на спеціалізованих корпусах значно покращує результати порівняно з загальною мовою. Це підтверджує важливість адаптації моделей до конкретної предметної області – у нашому випадку до українських наукових рефератів і англійських препринтів з ArXiv. Також слід зазначити монографію [5], де було запропоновано методи побудови тематичних графів на основі векторних представлень документів, що має пряме відношення до формування інтерактивної семантичної мережі в нашій системі.

Метою дослідження є підвищення ефективності пошуку та аналізу наукової інформації в реферативних базах даних шляхом інтеграції великої мовної моделі (LLM) у традиційну пошукову інфраструктуру. Це передбачає не заміну, а доповнення класичних методів семантичними технологіями, які забезпечують глибше розуміння змісту документів, автоматизоване формування узагальнень, переклад, а також побудову інтерактивних мереж знань.

Для опрацювання текстових даних використовується комбінація класичного пошукового механізму та сучасних методів машинного навчання. Основним пошуковим ядром є Manticore Search, який забезпечує швидке виконання запитів до індексованого тексту за допомогою логічних та регулярних виразів. У паралельному режимі працює система на основі LLM, яка забезпечує семантичне збагачення початкових даних.

Семантичне індексування на основі LLM

Кожен текст проходить попередню обробку, після чого подається на вхід LLM. На виході отримується числове представлення документа у вигляді вектора $v_d \in R^n$,

де n – розмірність прихованого простору моделі. Цей вектор кодує семантичну структуру тексту:

$$v_d = f_\theta(T_d),$$

де T_d – текст документа d ,

f_θ – функція, реалізована мережею з параметрами θ .

Вектори всіх документів зберігаються в окремому індексі, побудованому за допомогою FAISS (Facebook AI Similarity Search), що забезпечує швидкий пошук найближчих сусідів у багатовимірному просторі.

Пошук на основі семантики

При постановці запиту q , модель перетворює його у вектор v_q . Документи, найближчі до цього запиту, визначаються за косинусною подібністю:

$$\text{sim}(v_q, v_d) = \frac{v_q \cdot v_d}{\|v_q\| \cdot \|v_d\|}.$$

Цей показник дозволяє ранжувати документи не лише за наявністю слів, а й за їхнім значенням. Така модель значно перевершує традиційні методи типу TF-IDF або BM25, особливо у випадках, коли запит сформульовано на вищому рівні абстракції.

Формування дайджестів і аналітичних зведень

LLM також використовується для автоматичного формування дайджестів, які узагальнюють результати пошуку. Для множини документів $D_q = \{d_1, d_2, \dots, d_k\}$, відповідних запиту q , будується загальна тематична модель $\Phi(D_q)$, яка далі перетворюється у зрозуміле людиною повідомлення A_q :

$$A_q = g_\Phi(\Phi(D_q)),$$

де g_Φ – генератор тексту, навчений на наукових анотаціях.

Переклад і аналіз окремих документів

Модель може також перекладати документи на інші мови, зберігаючи їхню наукову точність. Для кожного тексту T_d формується його переклад T'_d на мову l :

$$T'_d = h_\Psi(T_d, l),$$

де h_Ψ — перекладацька функція, реалізована LLM.

Окрім того, проводиться детальний аналіз окремих документів, який включає виділення головної мети, методології, результатів та висновків.

Побудова інтерактивної семантичної мережі

На основі множини векторів $V = \{v_1, v_2, \dots, v_m\}$, відповідних певній темі, будується граф понять $G=(N, E)$, де N – вершини (поняття, автори, терміни), E – ребра, що вказують на взаємозв'язки. Кожне ребро має вагу $w_{i,j}$, яка обчислюється як подібність векторів:

$$w_{i,j} = \text{sim}(v_i, v_j)$$

Цей граф динамічно оновлюється та відображається у вигляді інтерактивної карти, доступної користувачеві.

Результатом інтеграції LLM стало створення нового покоління пошукового сервісу, який поєднує точність традиційного пошуку з глибиною семантичного аналізу. Серед основних досягнень:

Підвищення точності пошуку на 15% порівняно з традиційними системами;

Автоматичне формування наукових дайджестів, які зменшують час аналізу на 30–40%;

Переклад наукових текстів без втрати контекстуальної точності;

Візуалізація семантичних зв'язків у вигляді інтерактивної мережі, що виявляє приховані залежності між дослідженнями.

Висновки. Застосування великої мовної моделі у рамках реферативної наукової бази даних «Бібліотека Наукометрії» довело ефективність такого підходу для аналізу великих текстових масивів. Математично обґрунтована модель семантичного індексування, використання косинусної подібності, автоматичне формування дайджестів та побудова інтерактивної мережі понять – все це дозволяє говорити про створення нового типу інтелектуального пошукового сервісу.

Новизна дослідження полягає в тому, що вперше для української реферативної наукової бази даних «Бібліотека Наукометрії» було комплексно застосовано велику мовну модель LLama-4 у поєднанні з традиційним пошуковим двигуном Manticore Search. Це дозволило не лише підвищити точність пошуку на 15%, але й розширити аналітичні можливості системи: автоматично формувати наукові дайджести, перекладати тексти без втрати контексту, а також створювати інтерактивні семантичні мережі, які виявляють приховані зв'язки між темами, авторами та концепціями. Особливим етапом стало адаптування моделі до українськомовних наукових текстів, що відкриває шлях до подальшого розвитку українських цифрових гуманітарних наук. Час, необхідний для аналізу результатів пошуку, скоротився на 30–40% завдяки автоматичному формуванню наукових дайджестів і перекладу документів. Крім того, інтерактивні семантичні мережі надали користувачеві новий рівень взаємодії з інформацією, що особливо корисно для наукових оглядів, освітніх матеріалів та прогнозування розвитку тем.

Подальший розвиток передбачає застосування адаптивного навчання на основі зворотного зв'язку від користувачів та розширення мовної підтримки для українськомовних текстів.

Dmytro Lande,

ORCID <https://orcid.org/0000-0003-3945-1178>,

Doctor of Technical Sciences,

Head of the Department, Professor,

National Technical University "Ihor Sikorskyi Kyiv Polytechnic Institute",

Kyiv, Ukraine

e-mail: dwl@gmail.com

Andrii Kriuchyn,
ORCID <https://orcid.org/0000-0002-5063-4146>,
Corresponding Member of National Academy of Sciences of Ukraine,
Doctor of Technical Sciences, Professor,
Deputy Director,
Institute for Information Recording of National Academy of Sciences of Ukraine,
Kyiv, Ukraine
e-mail: kryuchyn@gmail.com

Application of a Large Language Model in the Technology of Traditional Search in Bibliographic Databases

This paper presents the integration of a large language model into the search system "Library of Scientometrics", which combines abstracts from Ukrainian scientific publications in the "Ukrainika Scientific" (1998–2024, 769 thousand records) and preprints from ArXiv (2007–2025, 2247 thousand documents). The core search engine is Manticore Search, enhanced with a semantic layer based on LLama-4 via API. A mathematical foundation for semantic indexing, cosine similarity-based search, and extended analysis modes – including digest generation, translation, and interactive concept networks – are proposed. The system's effectiveness in thematic search, trend detection, and author network construction is demonstrated. Search accuracy increased by 15%, while analysis time was reduced by 30–40%.

Keywords: Large language model, LLM, Manticore Search, semantic search, bibliographic database, digests, interactive semantic network, ArXiv, Ukrainika Scientific, scientometrics.

Ласенко Наталія Анатоліївна,
ORCID <https://orcid.org/0009-0004-3524-271X>,
заступниця директора,
бібліотека Національної музичної академії імені П. І. Чайковського,
Київ, Україна
e-mail: natalialaenko@gmail.com

Формування фонду бібліотеки Національної музичної академії імені П. І. Чайковського: специфіка та пріоритети

Розглядається специфіка та сучасний стан формування фонду музичною літературою в бібліотеці Національної музичної академії України імені П. І. Чайковського. Визначено проблемні питання, з якими стикається бібліотека в умовах обмеженого асортименту музичної літератури, запропонованого видавництвами України. Проаналізовано роль дарунків у формуванні бібліотечного фонду, а також окреслено потенційні шляхи для покращення процесів забезпечення освітньої та наукової діяльності сучасними фаховими виданнями.

Ключові слова: бібліотека, формування фонду, музична література, дарування, музична освіта

Бібліотека Національної музичної академії України імені П. І. Чайковського (НМАУ) є провідним інформаційним та науково-методичним центром для студентів, викладачів і науковців у сфері музичного мистецтва. Основною функцією бібліотеки є інформаційне забезпечення освітньої та наукової діяльності академії шляхом комплектування фонду сучасною літературою. Бібліотека НМАУ – це справжній культурно-музичний центр із глибокою історичною традицією та багатими колек-

DOI :

УДК 02:001:316.77]:027.021.1:021:021.64](100)(082)

Бібліотека. Наука. Комунікація. Пріоритети сьогодення та перспективи : матеріали Міжнар. наук. конф. (1–3 жовт. 2025 р.) : у 2 т. / НАН України, Нац. б-ка України ім. В. І. Вернадського, Асоц. б-к України ; відп. ред. О. М. Василенко. Київ, 2025. Т. 1. 528 с.

Збірник містить матеріали міжнародної наукової конференції «Бібліотека. Наука. Комунікація. Пріоритети сьогодення та перспективи» (1–3 жовт. 2025 р.), в яких висвітлено пріоритетні напрями та актуальні питання розвитку бібліотечно-інформаційного потенціалу, проблеми функціонування наукових бібліотек в умовах воєнного стану та шляхи їх подолання, представлено результати досліджень та науково-прикладні розробки в сфері бібліотечної, інформаційної та архівної справи.

Українськими та зарубіжними фахівцями розкрито теоретичні та прикладні аспекти розбудови, використання, організації доступу до електронних інформаційних ресурсів наукової бібліотеки у вебсередовищі та оптимізації бібліотечно-інформаційної діяльності задля продукування нового знання і забезпечення інформаційних потреб суспільства. Розглянуто актуальні питання щодо міжнародної взаємодії та сталого розвитку бібліотечно-інформаційного комплексу; питань відкритої науки в Україні; бібліотечно-інформаційної діяльності у сфері правового і соціально-політичного інформування; ролі бібліотеки в цифровій інфраструктурі знань; рукописної спадщини України у сьогоденні та перспективі; пошуку відповідей на виклики часу української біографіки; сьогоденних викликів і нових можливостей наукової архівної спадщини; друкованої та електронної періодики в системі документальних комунікацій; історії, сучасного стану і тенденцій розвитку наукової каталогізації стародруків і рідкісних видань; дослідження музичної культури в умовах сьогодення; наукового, освітнього та культурного простору під час війни; особливостей реагування на сучасні виклики та світовий контекст інституту президентства в Україні. Також висвітлено новітні завдання збереження бібліотечних та архівних фондів, що становлять культурне надбання України; дослідників української діаспори; дослідження єврейської історико-культурної спадщини України (до 10-річчя Української асоціації юдаїки).

Видання адресоване працівникам бібліотечно-інформаційної галузі, фахівцям документно-комунікаційної сфери, викладачам, аспірантам і студентам закладів вищої освіти.

Відповідальний редактор: О. М. Василенко

Редакційна колегія:

Л. А. Дубровіна – чл.-кор. НАН України, д-р іст. наук; Т. Д. Антонюк – д-р іст. наук; Н. П. Бондар – канд. іст. наук; О. А. Вакульчук – д-р іст. наук; О. М. Василенко – канд. іст. наук; Т. І. Власюк – канд. техн. наук; О. В. Григоревська – д-р наук із соц. комунікацій; В. В. Добровольська – д-р наук із соц. комунікацій; М. Б. Закіров – д-р політ. наук; Л. В. Івченко – канд. мист.; Г. В. Індиченко – канд. іст. наук; Ю. М. Капіца – д-р юр. наук.; Т. В. Коваль – канд. іст. наук; Т. М. Коваль – канд. іст. наук; Ю. С. Ковтанюк – канд. іст. наук; Ю. М. Половинчак – д-р наук із соц. комунікацій; В. І. Попик – чл.-кор. НАН України, д-р іст. наук; Г. Я. Рудий – д-р іст. наук; Я. І. Савчук – канд. біол. наук; О. П. Степченко – канд. іст. наук; В. М. Удовик – д-р наук із соц. комунікацій; В. В. Черноіваненко – канд. іст. наук; А. І. Шаповал – д-р іст. наук.

Автори несуть повну відповідальність за підбір та точність наведених фактів, цитат, власних імен, бібліографічних посилань та інших відомостей.

Гуренко Є. Ю. Проблеми організації тематичного пошуку в електронному каталозі Національної бібліотеки України імені В. І. Вернадського	264
Данильченко І. В. Специфіка організації електронних персональних колекцій	267
Деркач Н. В. Псевдоніми в стартових авторитетних записах у файлі «Індивідуальний автор» (ATHRA) в процесі ретроспективної конверсії колекцій Національної бібліотеки України імені В. І. Вернадського	270
Добра Н. В., Сосідко І. В., Ясінська О. А. Анотований тематичний навігатор ресурсами Інтернету як інструмент забезпечення оперативного доступу до інформації	273
Добровська С. В., Овсієнко Л. М. Представлення наукового напрямку «Воєнна наука. Військова справа» в реферативній базі даних «Україніка наукова»	276
Жабін О. І. Деякі історичні аспекти становлення вітчизняної наукометрії: від теоретичних засад до сучасних практичних рішень	280
Заїка В. М. Інтероперабельність електронних інформаційних ресурсів Національної бібліотеки України імені В. І. Вернадського: поточний стан	282
Ісаєва О. В., Дорош М. В., Власова Т. Ю. Авторитетна робота як модель пов'язаних даних: досвід Національної бібліотеки України імені В. І. Вернадського	286
Кириленко С. Е. Реферативні огляди та аналітичні доповіді в парадигмі інформаційного суспільства: контент-аналіз	289
Клюшнікова О. В., Сандул О. Г. Інформаційна діяльність бібліотеки в парадигмі суспільства знань	292
Колоскова Г. В. Кластерний підхід як форма організації бібліотечно-інформаційного потенціалу регіону в умовах цифровізації	295
Косенко О. В. Колекція А. Ю. Кримського у фондах Національної бібліотеки України імені В. І. Вернадського (на прикладі видання «Вісті Таврійського університету»)	298
Крючин А. А., Манько Д. Ю., Гарагуля С. С., Солоніна Н. В. Організація пошуку інформації в реферативних базах даних з використанням сучасних технологій ...	301
Кубко А. Ю. «Смартизація» бібліотеки: плюси і мінуси	304
Кудименко Л. Г. Research Organization Registry – новий стандарт у науковій комунікації	307
Ланде Д. В., Крючин А. А. Застосування великої мовної моделі у технології традиційного пошуку в реферативних базах даних	310
Ласенко Н. А. Формування фонду бібліотеки Національної музичної академії імені П. І. Чайковського: специфіка та пріоритети	314
Лощинська Н. В., Руденко Є. В. Нормативно-правова база розвитку бібліотечних електронних ресурсів в Україні	316
Міськов Г. В. Трансформація бібліотек України в умовах інтеграції у цифрову інфраструктуру знань: стратегічні виклики та інноваційні рішення	319
Островська О. М. База даних «Науковці України» – фундаментальний компонент інформаційного порталу «Наука України: доступ до знань» Національної бібліотеки України імені В. І. Вернадського	322