УДК 004.9
№ держреєстрації 0115U000422C
Інв. № _____

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Одеський національний політехнічний університет (ОНПУ)

65044, м. Одеса, пр.Шевченка, 1, тел. (048) 705 84 89

ЗАТВЕРДЖУЮ
Проректор

_____Д.В. Дмитришин
“____”_____2016 р.

З В І Т
З НАУКОВО–ДОСЛІДНОЇ РОБОТИ

Методологічні основи створення інформаційного середовища управління науковими дослідженнями структурних одиниць ВНЗ МОН України
№ 696 – 32

Технічна складова методології проектно-векторного управління інформаційними середовищами для моніторингу та управління науковими дослідженнями наукових груп
(проміжний)

Керівник НДР,
завідувач кафедри,
д-р техн. наук В.Д. Гогунський

2016

Рукопис закінчено 30 грудня 2016 р.

СПИСОК АВТОРІВ

Керівник НДР, головний науковий співробітник, д-р техн. наук, професор	В.Д. Гогунский (реферат, вступ, розд. 1 – 3, висновки)
Відповідальний виконавець, старший науковий співробітник, канд. техн. наук, доцент	О.Є. Колесніков (розд. 2, 4)
Старший науковий співробітник, д-р техн. наук, доцент	К.В. Колеснікова (розд. 1, 4)
Молодший науковий співробітник, канд. техн. наук, асистент,	А.С. Коляда (розд. 2, 4)
Старший науковий співробітник, канд. техн. наук, доцент	Т.М. Олех (розд. 2)
Молодший науковий співробітник, аспірант	А.О. Негри (розд. 4)
Молодший науковий співробітник, аспірант	В.О. Яковенко (розд. 1)
Молодший науковий співробітник, аспірант	В.Ю.Васильєва (розд. 1)
Молодший науковий співробітник, аспірант	Шерстюк О.І. (розд. 1)
Молодший науковий співробітник, асистент	Дмитренко К.М. (розд. 1)
Молодший науковий співробітник, інженер	О.М. Миколюк (розд. 3)

У виконанні окремих завдань приймали участь: Оборська Г.Г., Бабюк С.Н., Лук’янов Д.В., Лебідь В.В., Чернявський О.І., Отрадська Т.В.,

Р Е Ф Е Р А Т

Звіт з НДР: 123 с., 36 рис., 13 табл., 141 джерел.

Розвиток та оновлення знань в освітній сфері щодо забезпечення продуктивної роботи багатьох освітніх, науково-освітніх, адміністративних, науково-методичних і науково-дослідних організацій, задіяних у процесах підготовки висококваліфікованих фахівців і виконання наукових досліджень, потребують вирішення важливих завдань у площині визначення особливостей управління освітньою сферою, аналізу умов її функціонування та формалізації управлінських функцій. На часі є перехід від формальних концепцій адміністративного управління до застосування векторної парадигми управління з формалізацією управління інформаційними середовищами.

Об'єктом дослідження є процеси управління організаціями наукової сфери.

Предметом дослідження є технічна складова проектно-векторного управління інформаційними середовищами для моніторингу та управління науковими дослідженнями наукових груп.

У дослідженні виконана формалізація завдань розробки методології управління проектами в інформаційній сфері України та визначено критерії оцінки ефективності інформаційної системи оцінки наукової діяльності структурних підрозділів ВНЗ МОН України.

Результати дослідження спрямовані на розробку теоретичних засад нової інформаційної системи наукової діяльності ВНЗ МОН України, яка не має аналогів у вітчизняній та зарубіжній практиці. Проект орієнтований на розв'язання протиріч розвитку наукових досліджень в Україні, які породжені різноплановостю виконуваних наукових досліджень, що унеможливлює порівняння та оцінювання наукових результатів в контексті нового бачення та розуміння важливості Європейського вектору розвитку України.

УПРАВЛІННЯ, НАУКОВІ ПУБЛІКАЦІЇ, ВИЛУЧЕННЯ, ВЕБ-СТОРІНКИ, МОНІТОРИНГ, ПРАКТИКА, ОЦІНКА, ІНФОРМАЦІЙНЕ СЕРЕДОВИЩЕ.

З М І С Т

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

ПВП	Проектно-векторний простір
ППП	Проекти / програми / портфелі
IPMA	Міжнародна асоціація проектного менеджменту
ISO	Міжнародна організація по стандартизації
PDCA	Цикл Шухарта – Деминга, PDCA (Plan – Do – Check – Action) – «план, здійснення, перевірка, дія»
PMI	Інститут проектного менеджменту
PMBoK®	Project Management Body of Knowledge (Керівництво до зводу знань з управління проектами), розроблений PMI
P2M	Керівництво з управління проектами і програмами розвитку підприємств, розроблений PMAJ
ЖЦ	Життєвий цикл
ЗСП	Збалансована Система Показників
LSA	Латентно семантичний аналіз (Latent Semantic Analysis)
LDA	Латентне розміщення Дірихле (Latent Drichlet Allocation)
SVD	Сингулярна декомпозиція (singular value decomposition)
ПЗ	Програмне забезпечення
ПП	Портфель проектів
ППП	Проекти / Програми / Портфелі проектів
ТЕО	Техніко-економічне обґрунтування
ТЗ	Технічне завдання
УП	Управління проектами
БД	База даниxx
НМБД	Наукометрична база даних

ВСТУП

Розвиток інформаційних технологій з організації міжнародних наукометричних баз даних (НБД) і електронних бібліотек породжує нові можливості і завдання у сфері освітньої та наукової діяльності у вищій школі України. Одним з напрямів цієї діяльності є визначення узагальненої оцінки якості і результатів наукових досліджень окремого вченого, кафедри, університету та вищих навчальних закладів України в цілому [1 – 3]. Аналіз характеристик та основних властивостей НБД та індикаторів цитування наукових публікацій відображають широкий спектр цільового призначення НБД – від забезпечення суто інформаційних потреб науковців до всебічного аналізу публікаційної активності авторів [4].

Упродовж останніх років МОН України цілеспрямовано орієнтує публікаційну діяльність учених ВНЗ на входження у світове наукове співтовариство. Активність публікації наукових співробітників є одним з основних факторів, який враховується при визначенні світових рейтингів вищих навчальних закладів. При цьому НМБД є основними осередками трансформації знань і каналами подальшого застосування наукових результатів, як головної інформаційної та соціальної характеристики країни, університету, наукового колективу або окремого вченого [5].

Існуючі НМБД, як правило, орієнтовані на пошук публікацій тільки у своїх сховищах [6]. При цьому різні НМБД використовують свої специфічні форми інтерфейсу, що визначає необхідність обов’язкової особистісної участі науковців у пошуку публікацій в різних базах. Ці обставини породжують протиріччя між необхідністю інтегральної оцінки публікаційної активності авторів і відсутністю інформаційних технологій, які дозволяють виконувати інформаційно-пошукові операції в різних НМБД. Крім того, існує проблема багатоваріантного завдання атрибутів пошуку, у тому числі, різних варіантів написання прізвищ. В різних НМБД не використовуються моделі, методи та інструментальні методи визначення достовірності отриманої інформації, які засновані

на аналізі прихованих змінних для виявлення зв'язків в наборі назв публікацій, що дозволяє достовірно ідентифікувати публікації конкретних авторів. Тому розробка програмних продуктів для вилучення інформації з різних НБД є актуальною проблемою, що спрямована на розв’язання вказаних протиріч у галузі інформаційних технологій [7 – 9].

Об'єкт дослідження. Інформаційна технологія пошуку та вилучення контенту даних в інтернет додатках.

Предмет дослідження. Методи вилучення і обробки контенту метаданих публікацій в наукометричних базах даних

Методи дослідження. Для досягнення мети і вирішення завдань, поставлених у дослідженні, використовується теоретичний аналіз способів доступу до інформації з наукометричних баз. Метод моделювання використовується при побудові імовірнісних моделей з прихованими змінними (для ідентифікації публікацій конкретного автора).

Метою дослідження є розробка способу вилучення метаданих публікацій з наукометричних баз даних завдяки створенню програмного продукту, який реалізує спосіб добування інформації з наукометричних баз даних, а також надає програмний інтерфейс для використання іншими засобами цього функціоналу.

Для досягнення поставленої мети вирішені наступні завдання:

– аналіз найбільш відомих наукометричних баз даних на предмет доступу до метаданих публікацій та ідентифікація універсального способу;

– розробка інформаційної технології вилучення метаданих публікацій з наукометричних баз даних на основі веб інтерфейсу;

– розробка програмної системи автоматизації вилучення метаданих публікацій з найбільш поширених наукометричних баз даних;

– розробка алгоритму ідентифікації публікацій конкретного автора по набору ключових слів або близьких за тематикою;

– розробка графічного інтерфейсу користувача для управління пошуком публікацій, їх перегляду та аналізу.

1 ТЕОРЕТИЧНІ ОСНОВИ РОЗРОБКИ ТА ЗАСТОСУВАННЯ ІНСТРУМЕНТАРІЮ ПЛАНУВАННЯ НАУКОВИХ ДОСЛІДЖЕНЬ

1.1 Формування цінності в проектно-орієнтованій діяльності

У сфері наукометричних вимірювань використовуються інформаційні об'єкти, які є наукометричними базами даних (НМБД) [10]. Вони являють собою електронні засоби збереження та обробки наукометричних показників, а також, найчастіше, містять тіло публікаційних матеріалів (статей, журналів, книг). Найбільш авторитетні НМБД світового рівня (Scopus, Web of Science, Springer) є реферативними базами даних, які включають в себе деяку кінцеву множину публікацій, а також засоби сервісу для задоволення інформаційних потреб користувачів. При цьому, за деяким винятком, надання інформаційних послуг здійснюється на платній основі [11].

Наукометричні показники поділяють на такі категорії: на основі числа публікацій (сумарне число, зважена сума, відношення кількості публікацій до наукового стажу); на основі кількості цитуваннь (сумарне число посилань, кількість прихованого цитування); на основі числа публікацій та кількості цитування (індекс Гірша і різні його модифікації) [16].

Міжнародна практика наукометричних досліджень сьогодні найбільш часто базується на використанні двох баз даних: Web of Science і Scopus [4]. Широко відомі також інші бази даних, які орієнтовані на інформаційне забезпечення наукових досліджень без формування даних наукометрії. Всі вони в основному не є комерційними базами. Серед некомерційних НМБД, в яких індексуються публікації українських вчених, можна назвати наступні: Copernicus, BASE, DOAJ, Driver, Science Index, UlrichsWeb та ін.

Останнім часом набирають популярність наукові соціальні мережі: Google Search, ResearchGate, Academia.edu, Mendeley. Ці безкоштовні програми орієнтовані на упорядкування та управління бібліографічною інформацією окремих науковців або певних структурних одиниць навчальних закладів і наукових ус-

танов. Базові пакети вказаних програмних засобів розповсюджується безкоштовно, проте існують платні версії зі збільшеними квотами на зберігання матеріалів і створення груп.

У більшості випадків, НМБД не містять в собі повного тексту наукових публікацій, а тільки метадані про неї і посилання на вихідний документ. Метадані – це дані про дані або структуровані дані, що представляють собою характеристики описуваних сутностей для цілей їх ідентифікації, пошуку, оцінки, управління ними. Відомі способи стандартизації метаданих публікацій для полегшення можливої обробки їх автоматизованими засобами. Одним з них є використання спеціальних репозиторіїв, які призначені для документообігу певного типу:

– Eprints – пакет вільного програмного забезпечення для побудови архівів відкритого доступу і в основному використовується для створення колективних архівів і наукових журналів;

– Dspace – вільна платформа для інституційних репозиторіїв (для довгострокового зберігання цифрових матеріалів – публікацій);

– Digital Commons – є відкритим інституціональним репозиторієм і видавничим рішенням, що поєднує традиційну функціональність з інструментами для рецензованих публікацій журналу;

– OJS (Open Journal Systems) – система призначена для створення електронних журналів з відкритим доступом і дозволяє організовувати весь робочий процес видання: прийом, рецензування та каталогізація статей.

Одним з глобальних трендів розвитку конкурентоспроможності підприємств і організацій, у тому числі і ВНЗ, особливо в умовах кризи, є перехід до проектно-орієнтованої діяльності, яка за визначенням Родні Дж. Тернера спрямована на управління змінами при реалізації проектів / програм / портфеля проектів (ППП) [17]. У сучасній культурі управління проектами (УП) формуванню цінності проектів на фазі їх ініціації приділяється недостатньо уваги, тому часто необхідно вирішувати протиріччя між оточенням, що безперервно змінюється, і цілями проектної діяльності, яка передбачає управління змінами та постій-

не удосконалення процесів і продуктів проектів на основі врахування найкращої практики і теорії проектного управління [18]. При цьому розвиток і активне застосування ціннісного підходу в проектно-орієнтованій діяльності організацій часто стримується через відсутність методів комплексної (багатофакторної) оцінки результативності ППП в динаміці життєвого циклу [19].

Імплементація нового Закону Україні «Про вищу освіту», а також прийняття нового Закону "Про наукову і науково-технічну діяльність" визначають, що процес інтеграції науки та освіти – одне з головних завдань реформування вищої освіти. Управління змінами в освітній галузі спрямовано на підвищення якості вищої освіти в Україні через проекти і заходи організаційно-технічного та наукового змісту [20 – 23]. Наразі, виходячи з концепції ціннісного підходу можна визначити загальний профіль цінностей проекту “Якість освіти в Україні” (рис. 1.1), що дозволить сформувати критерії оцінки якості освітньої діяльності, зокрема наукових здобутків, які відображають наукову складову діяльності вищих навчальних закладів (ВНЗ) через наукові дослідження, публікації, впровадження у навчальний процес і бізнес нових технологій.

Рисунок 1.1 – Профіль цінностей проектно-орієнтованої діяльності за напрямом “Якість освіти в Україні”

Зазначені на рис. 1.1 складові орієнтують ВНЗ України на вихід на міжнародний рівень через наукові дослідження і публікації отриманих результатів у міжнародних наукових виданнях [24].

Зростанням вимог до теоретичної і практичної спрямованості наукових досліджень обумовлює необхідність ефективного використання сучасних інформаційних технологій та методів управління науковими дослідженнями в масштабах кафедри, факультету, ВНЗ та освітньої галузі в цілому. При цьому одним з основних показників ефективності наукових досягнень є наукові публікації. Саме множина публікацій становить основу формування нових знань щодо розширення можливостей результатів досліджень та створення нової цінності в світовій економіці. Оскільки теоретичні, функціональні і структурні зміни в різних галузях знань в певній мірі відображаються у наукових публікаціях [25].

Нині публікаційна активність з особистої зацікавленості науковців трансформована у реальні показники діяльності ВНЗ. Як прийнятний чинник оцінювання діяльності ВНЗ часто розглядається показник «чисельність науково-педагогічних працівників, які мають публікації у виданнях іноземних держав або у виданнях України, які включені в міжнародні наукометричні бази у звітному навчальному році» [26]. Державні вимоги з акредитації ВНЗ містять показники числа публікацій і цитування науковців у виданнях, які входять до міжнародних науково метричних баз.

Головним фактором зміни зовнішнього середовища є наростання конкуренції між ВНЗ за обмежені ресурси, оскільки щорічно через демографічний спад кількість випускників середніх шкіл скорочується. В останні роки змінюється відношення суспільства до праці в освітній сфері, тому за відсутністю мотивації склад викладачів вищої школи поповнюється недостатньо. Кризові явища в економіці призводять до зменшення попиту на підготовку фахівців, що в свою чергу, веде до скорочення можливостей фінансової підтримки освітніх закладів. Сьогодні, коли основні ніші вже заповнені і кон'юнктура попиту істотно модифікувалася у напрямку спеціального реального сектора економіки, необхідно прояснення концепції розвитку ВНЗ, аналіз їх конкретних переваг і визначення

тієї стратегії, яка забезпечить стабільність розвитку, будь то стратегія диверсифікації, поглиблення спеціалізації або ін.

За цих обставин проектно-орієнтоване управління ВНЗ стає нагальною необхідністю для розробки та прийняття довгострокових заходів щодо формування і підтримки конкурентних переваг на ринку послуг вищої освіти. При цьому найчастіше зусилля керівників вищих навчальних закладів спрямовуються на виконання окремих стратегічних програм, що ґрунтуються на прагненні досягнення реальних результатів в одному з пріоритетних напрямів – удосконаленні процесів управління ВНЗ [27].

У сучасній системі освіти має місце ситуація, коли всі причетні до управління освітньою діяльністю ВНЗ, ратують за підвищення якості цього процесу, але при цьому кожна категорія оцінює це поняття по-своєму. Така невідповідність призводить до неузгодженості в роботі органів управління освітою та освітніх установ, і не сприяє досягненню головної мети щодо забезпечення якості освіти. Існуюча система оцінки результатів навчальної роботи центральними органами управління освітою, не спрямовує ВНЗ до впровадження нових принципів та механізмів досягнення і безперервного поліпшення якості навчальної роботи, а лише фіксує певні і не завжди достовірні показники, без аналізу можливих причин, що сприяли їх формуванню [28].

Центральні органи управління освітою не завжди здатні врахувати специфічні особливості того регіону, в якому функціонує той чи інший ВНЗ, і детально вивчити його профільну орієнтацію. Більш того, в умовах ринкових відносин у сфері освіти, на центральні органи управління освітою такі завдання й не покладені. Центр, як правило, обмежується завданням загальних вимог (стандартів) до освітньої системи, з наступним контролем їх дотримання шляхом проведення акредитації, атестації, планових перевірок та інших заходів. У цих умовах навчальний заклад не може виступати в пасивній ролі «статиста», а повинен надавати активний вплив на хід навчального процесу з метою його безперервного поліпшення з урахуванням регіональних особливостей і домінуючого профілю. Іншими словами, ВНЗ повинен розглядатися як активний компо-

нент в загальній системі освіти, а в системі управління закладом повинні реалізовуватись механізми активної самоорганізації, що реалізуються в рамках зовнішніх обмежень (освітніх стандартів). В умовах ринкових відносин у сфері освіти ВНЗ слід розглядати як складну відкриту систему, що самоорганізується, здатну забезпечувати якісну освіту в ситуаціях конкуренції за рахунок збереження системного гомеостазу і оперативного пристосування до динамічних умов зовнішнього середовища [29].

Дана модель дозволяє синтезувати і визначати оптимальну структуру та повноваження підрозділів ВНЗ в єдиній інтегрованій системі для досягнення необхідних показників якості навчальної роботи. Система управління якістю навчальної роботи має будуватись на профілях створюваної цінності для освітніх проектів. Концепція проектного управління освітніми закладами на засадах створення цінності дозволяє перейти від одномірного до багатовимірного оцінювання проектів. Такий підхід є суттєвим для проектів освітньої спрямованості. Для цих проектів слід ураховувати множину факторів зовнішнього оточення, потреби суспільства, властивості створюваного продукту, характеристики і рівень досконалості процесів, тенденції розвитку ВНЗ. Далі під продуктом освітніх проектів будемо розуміти новий стан, у який сукупність випускників ВНЗ переведена у наслідок виконання освітніх проектів [30]. Тобто продуктом освіти є випускники з новими знаннями, навичками та уміннями, що формують необхідні для фахівців виробничі та суспільно значимі компетенції. Створювана цінність в освіті може бути відображена як кортеж:

$$\mathbf{C = \{(вид\ цінності \leftrightarrow драйвери \leftrightarrow засоби \leftrightarrow показники)i \leftrightarrow індикатори\},}$$

де $i = 1, 2, \ldots, n$ індекс виду цінності освітніх проектів.

При цьому індикатори є оцінкою досягнутого рівня досконалості за певним видом цінності, що характеризують інтегральну оцінку проекту. Так, ефективність проекту залежить від цінностей продукту, процесу, діяльності, а також цінності розвитку і оновлення (рис. 1.1).

Модель профілювання цінності в проектах освітньої спрямованості дозволяє перейти від одновимірної оцінки ефективності складних навчальних систем до багатовекторного оцінювання за множиною характерних параметрів. У структурі оцінки створюваної цінності запропоновані драйвери інноваційного розвитку та засоби, що відображають у чіткій або нечіткій формі результати діяльності ВНЗ. Розроблена модель системи управління якістю навчальної роботи ВНЗ на засадах створюваною цінності, дозволяє оцінити можливості успішного виконання функціональних завдань керівників різного рівня з управління навчальним процесом для досягнення заданого рівня якості діяльності ВНЗ за показниками наукової цінності освітніх проектів.

Світова спільнота вчених вже давно оцінює результативність наукової діяльності використовуючи кількісні показники. Кількісні оцінки ґрунтуються на оцінці числа публікацій, частоті їх цитування, застосуванні загальноприйнятих наукометричних показників, таких як індекс Гірша (*h*-індекс) або імпакт-фактор наукового журналу. З перерахованих показників останнім часом чи не найпоказовішим можна вважати індекс Гірша [31]. Адже він є кількісною характеристикою якості та продуктивності роботи вченого, групи вчених, університету або країни в цілому, і визначається на основі кількості публікацій і кількості цитувань цих публікацій.

Конкуренція у сфері вищої освіти породжує створення нових механізмів управління науковими дослідженнями, що спонукає наукові колективи і окремих науковців до аналізу своєї публікаційної активності для пошуку активних заходів щодо покращення показників цитування публікацій [32]. При цьому науковий внесок в розвиток теорії і практики, що міститься у наукових статтях, запропоновано визначати на основі показників цитування статей. Дійсно, цитування колегами певних статей у своїх публікаціях є оцінкою, як правило, позитивною статей, що цитуються. Наявність множини доступних наукометричних баз, різних пошукових систем і соціальних мереж науковців у світовій павутині створюють умови для діяльності щодо покращення показників цитування [33]. Адже важко спростувати очевидний факт, що цитованими є такі публікації, які є

доступними широкому загалу науковців, які є прочитаними і які містять незаперечну новизну або практичну цінність. Тобто для того, щоб певна стаття була цитованою, необхідно, аби вона була прочитана якомога більшою кількістю фахівців і науковців [34].

1.2 Життєвий цикл публікацій.

Зазвичай на основі виконаних експериментальних або теоретичних досліджень автори готують статтю до публікації. Редакції журналів редагують статті, направляють їх на рецензування [35]. Після позитивної рецензії статті готового примірника журналу розміщуються редакцією у різних депозитаріях, а також у НДБ, у яких індексується наукове видання (рис. 1.2).

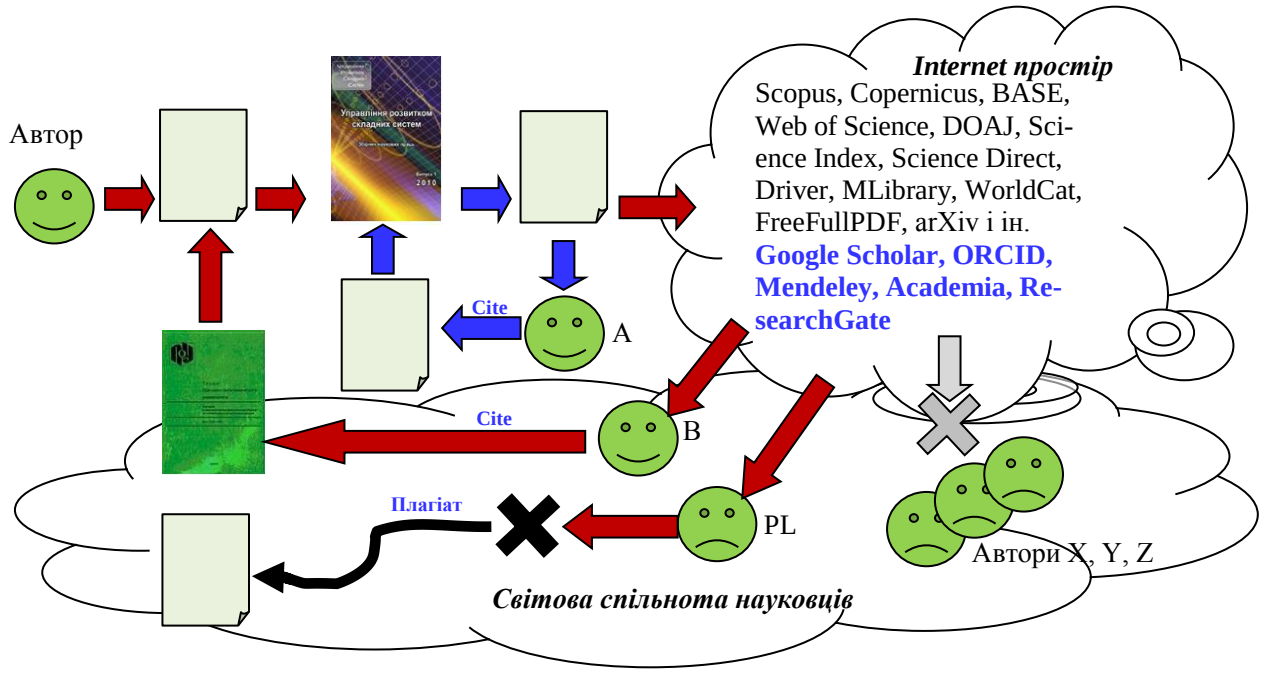


Рисунок 1.2 – Життєвий цикл публікації - ближній і дальній шлях цитування

Далі починається «самостійне життя» публікації. Наукова спільнота (A, B, …) отримує можливість ознайомитись зі змістом статті, пошукові автомати НДБ вилучають метадані статей (автори, організація, анотації, пристатейний список літератури). Метадані використовуються для визначення показників цитування.

Як показано схематично на рис. 1.2, об’єктивно існують ближній і дальній шляхи (цикли) цитування публікацій. Ближній цикл пов'язаний з появою посилання на публікацію у тому ж журналі, де була опублікована стаття. Дальній цикл – цитування виконується у іншому журналі. Разом з тим існує певна ймовірність, що деякі автори (PL) можуть запозичити частку матеріалу статті без посилання на першоджерело. Крім того слід зазначити, що деякі науковці (X, Y, Z) взагалі не отримують доступ до публікації через різні причини.

Зазначені особливості життєвого циклу публікацій породжують просте питання: «У який спосіб можна збільшити показники цитування?» Слід зазначити, що автори публікації, як було вказана вище, на цьому етапі життєвого циклу статті є відстороненими і не можуть активно впливати на те, щоб їхню роботу цитували інші автори. Тому базуючись на схемі рис. 1.2 можна зробити основну рекомендацію, що статті слід публікувати у фахових виданнях, де колеги зможуть ознайомитись зі статтею і оцінять її позитивно шляхом цитування [36].

Принципова схема управління процесом, що показана на рис. 1.3, містить цикл управління, у якому спільнота авторів або окремі науковці самі обирають засоби {*A, B, C ... Y, Z*} для розповсюдження результатів своїх досліджень у журналах, репозиторіях або у комунікаційних Internet – системах. Таким чином, розміщення публікацій слід віднести до елементів управління системою.

Разом з тим, як видно з рис. 1.2, існує і інший шлях просування публікацій до читачів у Інтернет-просторі. Цей шлях пов'язаний з активною участю авторів статей у розміщенні своїх публікацій у таких інформаційних системах, як Google Scholar, ORCID, Mendeley, Academia, ResearchGate [6]. Звісно, що ведення множини своїх публікацій у цих системах є досить затратним з точки зору витрат часу. Але, на нашу думку, такий підхід є виправденим – ніхто окрім автора не може об’єктивно представити наукові результати.

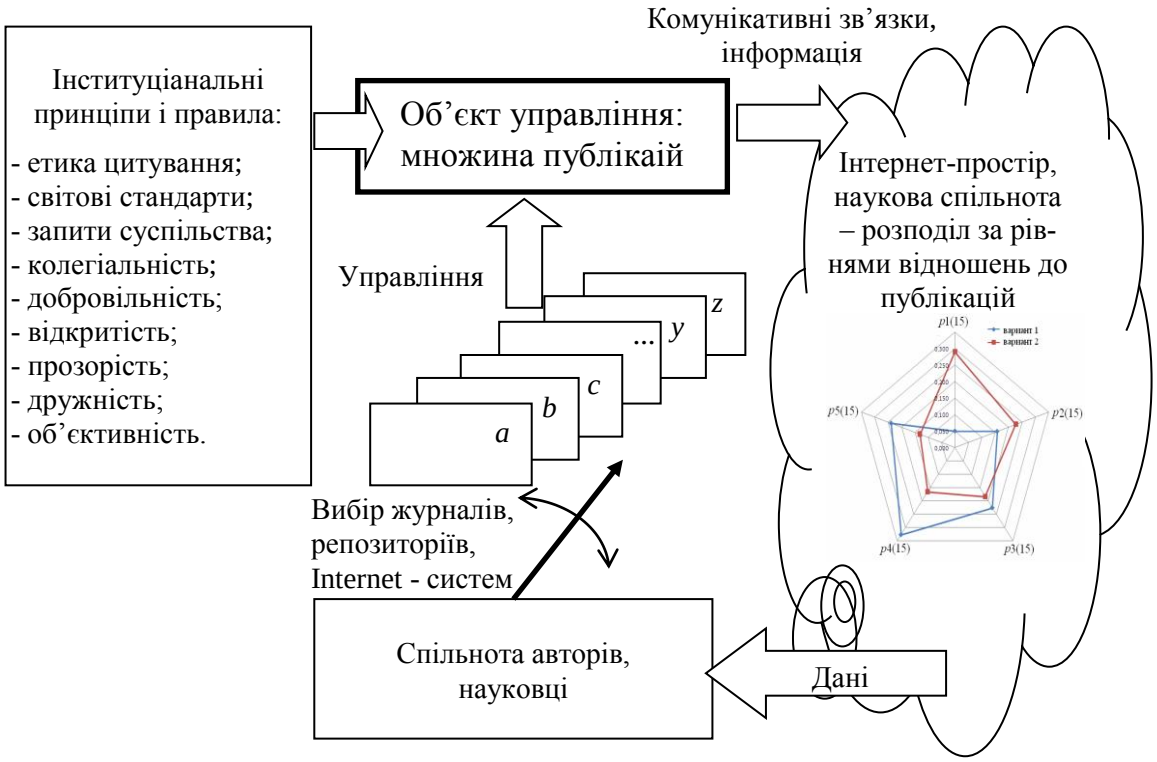


Рисунок 1.3 – Принципова схема управління процесом

Для сучасного стану наукометричних досліджень характерними рисами є формування умов автоматизації процесів пошуку статей [96]. Особливо важливим це є в наукоій сфері. Природно, що ця задача не може вирішуватися без знань основних закономірностей наукових комунікацій, без освоєння методів об'єктивного і своєчасного контролю й моделювання станів системи науковців, без технічних засобів використання цієї інформації для управління процесами [37 – 41].

1.3 Управління проектами підвищення активності публікацій в інформаційних інтернет-ресурсах

Метод ідентифікації авторів, заснований на визначенні потенційних авторів з написання статей, з урахуванням наукових колективів і частоти їх появи, дозволив підвищити якість визначення та зв'язку авторських профілів з наукометричними базами даних і користувачами інформаційно-пошукової системи [42]. Це є основою для визначення відповідності між користувачами інформа-

ційно-пошукової системи і їх профілями в відкритих наукових Інтернет-ресурсах [43].

Інформаційна модель профілю публікації, представлена на рис. 1.4.

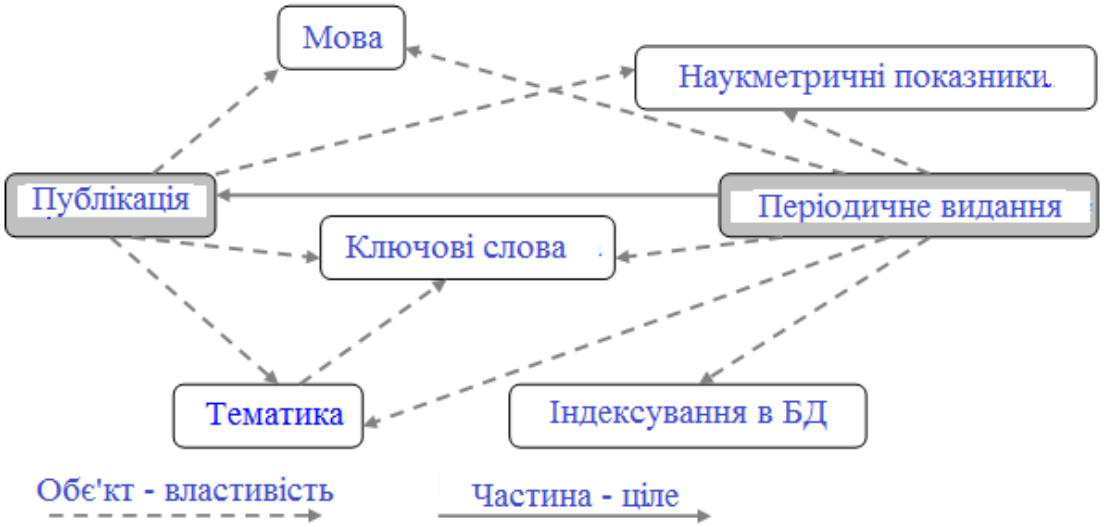


Рисунок 1.4 – Сутності інформаційної моделі публікації

Одним з показників профілю публікації є ключові слова K^s:

$$K^s = \bigcup_{i=1}^{l} K_i^s,$$

де l – число джерел ключових слів для періодичного видання.

Множина K^s ключових слів публікацій формуються на основі:

– безлічі ключових слів, зазначених авторами; безлічі ключових слів, отриманих з наукометричних баз даних;
– безлічі ключових слів періодичного видання, в якому опублікована стаття.

Зазначені безлічі ключових слів утворюють інформаційне поле області знань. Показники активністю публікацій організації та окремих учених набули статусу індикаторів затребуваності результатів наукових досліджень [44]. Тому потрібно визначення структури активністю публікацій, її впливу на проекти організації і вибір ефективних заходів управління.

Вищим результатом управління публікаційною активністю є досягнення таких значень показників, при яких можливо просування наукових розробок на внутрішній і зовнішній ринок, а також підвищення високих положень в рейтингових системах [45].

Структура активності публікацій ВНЗ України містить три основні компоненти: видавничий проект, публікації в українських видавництвах різного статусу, публікації в зарубіжних видавництвах. Кожен компонент цієї структури будується типовим чином і включає в себе: публікацію статей в журналах і збірниках, публікацію тез і доповідей за результатами роботи конференції. Формування наукових профілів конкурсів і наукових заходів проводиться за схожою схемою.

Для просування публікацій і підвищення показників активністю публікацій необхідно розвиток системи управління. Функції управління процесом активністю публікацій в рамках видавничого проекту пов'язані з аналізом результатів, моніторингом діяльності, урахуванням публікацій, інформаційних супровід процесів і стимулюванням авторів. Поєднання всіх складових процесу управління публікаційній активністю дозволяє отримати достовірну інформацію про протікання процесу і прийняти ефективне рішення про реалізацію заходів, спланованих у попередні і подальші періоди часу.

Важливість і значимість визначення цілей для проектно-орієнтованих організацій, що ініціюють і виконують проекти, очевидна з точки зору зрілості бізнесу [46]. Якісне цілепокладання системно знижує розпорошення ресурсів, консолідує зусилля всіх співробітників підприємства на шляху реалізації місії підприємства. Узгодженість особистих цілей співробітників, цілей бізнес-процесів, проектів, цілей підприємства або організації є необхідною умовою ефективності роботи підприємства, а управлінські навички цілепокладання є однією з найважливіших характеристик управління ППП [47].

1.4 Приклад моніторингу публікаційної активності науковців та кафедр вищих навчальних закладів в Google Академія

Проблема оцінки наукових досягнень окремого науковця і наукових колективів є однією з найактуальніших і в той же час найскладніших проблем, що стосуються взаємин, як усередині самої науки, так і з суспільством [13]. Можливі різні варіанти оцінки будь-якого виду творчої діяльності; проте у всіх сферах найбільш об'єктивною є оцінка за кінцевим результатом, а не за процедурою його досягнення і витраченим на це зусиллям [48].

В ідеалі засоби об'єктивної оцінки наукової діяльності повинні чимось нагадувати процедуру виявлення переможців у спортивних змаганнях, коли кращих визначають, орієнтуючись на ті чи інші досягнуті кількісні показники.

В останні десятиліття якісні критерії оцінки наукової діяльності представляються вже недостатніми і нагальною вимогою часу стає необхідність використання кількісних параметрів, які характеризують наукову діяльність і не залежать від будь-яких суб'єктивних факторів. Особливої важливості набуває подібна об'єктивна оцінка, коли мова йде про ті чи інші «відзнаки» окремого вченого або наукового колективу, фінансуванні наукових досліджень у вигляді грантової підтримки або заохочення окремих дослідників у вигляді присудження їм премій, медалей, ступенів та звань. Проте зараз, по суті, немає об'єктивних кількісних критеріїв оцінки наукової діяльності, хоча у світовій спільноті науковців є окремі дослідження і пропозиції щодо переваг формальних підходів до оцінки результативності наукової праці.

Для активізації науковців ВНЗ щодо публікацій результатів своїх досліджень у зарубіжних журналах або у виданнях України, що включені до зарубіжних наукометричих баз, Міністерство освіти і науки України запроваджує низку заходів [49 – 53]. Нові вимоги до наукових публікацій та безпосереднє оцінювання ВНЗ за числом публікацій, які індексовані у іноземних наукометричних базах даних (у першу чергу Scopus!) трансформують публікаційну активність з особистої зацікавленості науковців у справу державного значення щодо формування іміджу України у царині міжнародних наукових зв’язків і ство-

рення сприятливих умов фінансування наукових досліджень за міжнародними грантами за участю науковців України [54]. Тому актуальним завданням є моніторинг публікаційної активності науковців ВНЗ України у виданнях, які індексуются міжнародними науко метричними базами даних (НБД).

Наукометричні бази даних – це бібліографічні і реферативні бази даних з інструментами для відстеження цитування статей, опублікованих у наукових виданнях. Найбільш відомими на сьогоднішній день є бібліографічні бази даних Scopus, Web of Science (Web of Knowledge), Astrophysics, PubMed, Mathematics, ChemicalAbstracts, Springer, Agris, GeoRef та ін. Широко застосовуються також відомі МНБД: Begell House Inc., Pleiades Publishing, Kluwer та інші. Всі вони є комерційними базами.

Серед некомерційних наукометричних баз з технічних наук можна назвати такі [35]: Science Direct, Copernicus, Science Index, DOAJ, BASE, Driver, MLibrary, WorldCat, FreeFullPDF, arXiv, Google Академія та ін.

Наукометричні бази не є чимось новим у науковому світі. Їх історія бере свій початок з 1870-х років, коли вперше з'явилися два індекси наукового цитування − індекс юридичних документів ShepardsCitations (1873 рік) та індекс наукових публікацій з медицини IndexMedicus (1879 рік), який існував аж до 2004 року. З розвитком інтернет-технологій з'явилися Web of Science, Scopus та інші наукометричні бази даних, а також Академія Google (GoogleScholar), яку теж можна віднести до міжнародних наукометричних баз.

Наукометричні бази даних в даний час виконують функцію авторитетних джерел бібліографічної інформації по науковій періодиці країни по конкретному коду спеціальності. Крім того, бібліографічні та реферативні бази даних є інструментом для відстеження цитування статей, опублікованих в тих чи інших наукових виданнях. Такий моніторинг дає можливість формувати рейтинги журналів в базах. Високий рейтинг журналу в наукометричній базі означає його затребуваність науковим співтовариством.

Проблемою оцінки вчених займається наукометрія − розділ наукознавства, що займається статистичними дослідженнями структури і динаміки наукової

інформації. Основними параметрами, що характеризують рівень цитування автора, є такі [12]:

1. Індекс цитування, кумулятивний індекс цитування − загальна кількість посилань на всі роботи автора за весь час його діяльності.

2. Імпакт-фактор (класичний, синхронний, Гарфільдівскій).

3. Індекс Гірша, *h*-індекс.

4. Інші індекси: Egghe's (g-index), Zhang's (e-index), Contemporary h-index (hc-index), AW-index, Multi-authored h-index, hInorm, hIannual.

Індекс цитування має подвійне тлумачення [36]. В Україні це поняття часто визначає число цитувань публікацій або відношення числа цитувань до базових показників публікацій – числа журналів, авторів та ін.

Сучасне тлумачення індексу цитування пов’язане з англомовною калькою цього поняття [3]. Під індексом цитування розуміється реферативна база даних наукових публікацій, що індексує посилання, зазначені в пристатейних списках цих публікацій і яка надає кількісні показники посилань (такі, як сумарний обсяг цитування, індекс Гірша та ін). З статей у журналах, що включені у реферативну базу, витягуються традиційна бібліографічна інформація (вихідні дані) і списки цитованої літератури (пристатейна бібліографія).

Показники цитування можна піддавати критиці, як показник, статистично недостовірний, що залежить від галузі знань (у біологів і медиків більше, ніж у фізиків, а у фізиків, відповідно, більше, ніж у математиків), від сумарної кількості фахівців з того чи іншого розділу науки , від поточної популярності дослідження, від географії журнальних публікацій, віку дослідника, тощо. Але, на жаль, зараз поки що не існує інших показників, які більш адекватно відображають результативність роботи вчених.

Індекс Гірша (*h*-індекс) є кількісною характеристикою продуктивності вченого, групи вчених, університету або країни в цілому, заснованою на кількості публікацій та кількості цитувань цих публікацій [39]. Розраховується цей показник таким чином: вчений має індекс *h*, якщо *h* з його *Np* статей цитуються як мінімум *h* раз кожна, в той час як решта (*Np* -*h*) статей цитуються не більше,

ніж *h* разів кожна. Іншими словами, учений з індексом *h* опублікував *h* статей, на кожну з яких послалися як мінімум *h* разів.

Визначення публікаційної активності кафедр доцільно також визначати з використанням програми пошуку *Publish orPerish.*

Publish orPerish − програма, яка виконує пошук і аналізує цитування публікацій. Під час виконання досліджень дуже зручно перевіряти, чи були даний текст або автор вже процитовані, як часто і де [33]. Якщо є інтернет-з'єднання, то за допомогою Publish or Perish можна отримати майже миттєво таку інформацію. Ця програма використовує запити Google Scholar. Пошук може виконуватися за прізвищем автора, за назвою видання, по групі слів або за певною фразою. Publish orPerish може визначити посилання (URL) на документ, де можна виявити потрібний текст у своєму запиті, загальну кількість цитат, які відповідають запиту, середню кількість посилань. Можливий пошук статей за прізвищем автора, за назвою журналу, розгорнутий пошук.

Приклад відображення результатів запиту у Publish orPerish за прізвищем «Гогунський OR Gogunsky OR Гогунский» показаний на рис. 1.5.

Publish or Perish розроблена у Мельбурнському університеті (Австралія) і використовує базу даних Google Scholar (Google Академія). За допомогою цієї програми можна виконати:

1. Аналіз цитувань автора «Authorimpact» (рис. 1.5);

2. Аналіз цитувань журналу «Journalimpact»;

3. Розширений аналіз цитування автора «Generalcitations».

Послідовність дій для проведення аналізу цитувань автора є така:

1. Запустити програму Publish orPerish.

2. У горизонтальному меню вибрати вкладку «Authorimpact».

3. У полі «Author's name» ввести всі можливі варіанти написання прізвища та ініціалів автора кирилицею та латиницею. Різні варіанти написання прізвища та ініціалів слід об’єднати логічним оператором OR. Кожний варіант написання прізвища та ініціалів треба взяти в лапки без крапок в ініціалах. Наприклад: «Гогунський ВД» OR «Gogunsky VD».

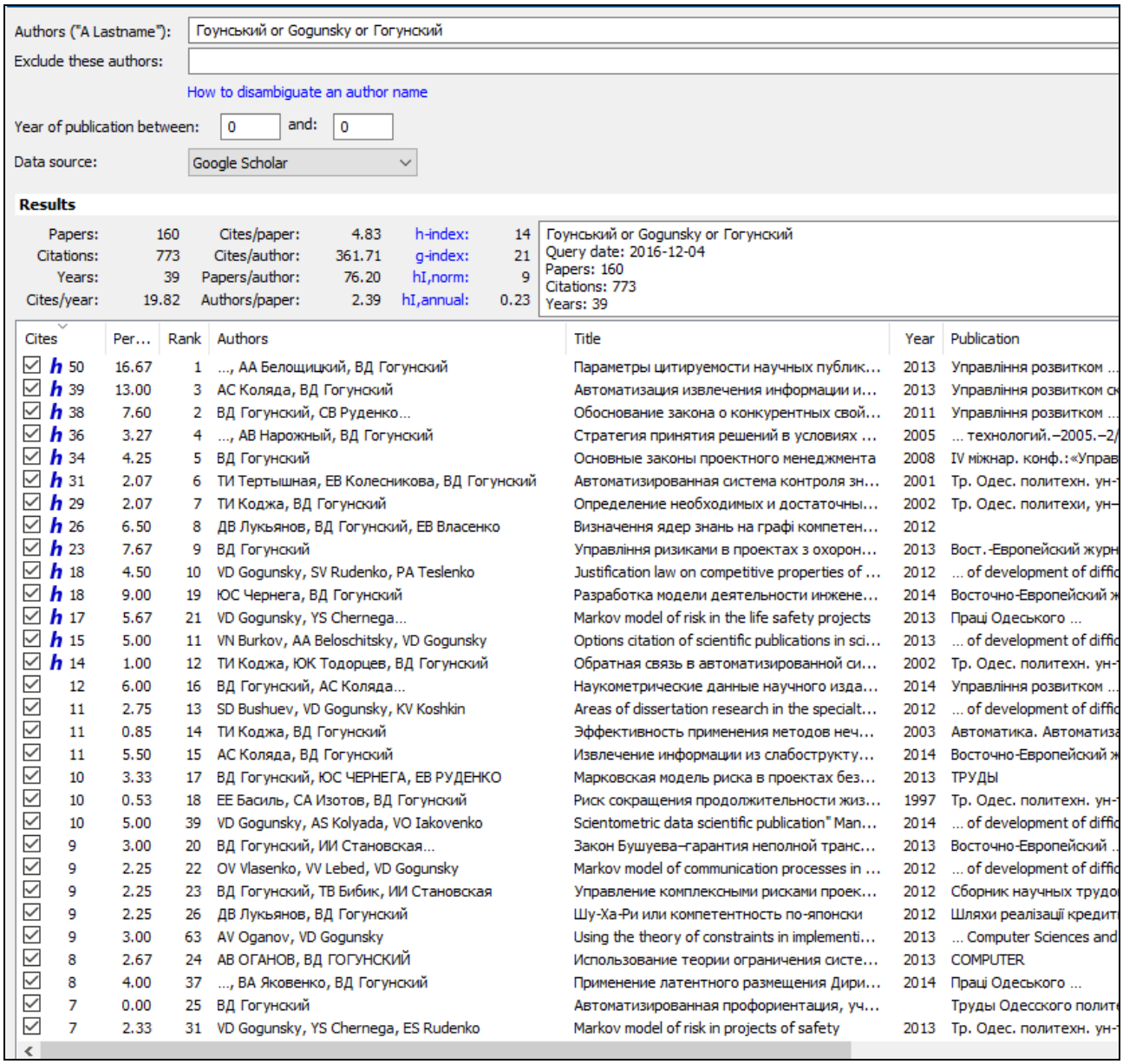


Рисунок 1.5 – Приклад відображення результатів запиту у Publish orPerish

4. Пошук при необхідності можна обмежити роками та тематичними напрямками.

5. Виконати команду Lookup після заповнення потрібних полів.

6. В головному вікні з’являються результати пошуку (рис. 1). Слід зазначити, що кількість знайдених результатів обмежена до 1000 записів, що обумовлено властивостями GoogleScholar.

7. В панелі статистичної інформації з’являється така інформація:

− загальна кількість документів автора;

− загальна кількість цитувань автора;

− середня кількість цитувань автора за рік;

− *h*-індекс (індекс Гірша);

− інші показники.

8. В панелі знайдених публікацій відображаються всі результати запиту. Список розбитий на такі колонки:

Cites – кількість цитувань конкретної статті;

Peryear – середня кількість цитувань статті за рік;

Rank – рейтинг статті Google Scholar;

Authors – всі автори статті;

Title – назва публікації;

Year – рік публікації;

Publication – назва журналу (в деяких випадках не визначається);

Publisher – видавець (в деяких випадках не визначається).

Якщо публікація має цитування (друга колонка, рис. 1), то подвійний клік на вибраному рядку веде на сформований Google Scholar список статей, які її цитують. Якщо цитувань немає, то відображається сторінка з результатами пошуку Google Scholar даної публікації.

9. Список публікацій за замовченням відсортовується у порядку зменшення кількості цитувань, але можна відсортувати його за значеннями будь-якого стовпця, натиснувши на відповідний заголовок.

10. Якщо окремий рядок не відповідає пошуковому запиту, можна виключити його з розгляду, знявши «галочку».

11. При потребі можна здійснити редагування шляхом об’єднання окремих рядків (якщо вони відносяться до однієї роботи). При цих змінах статистичні показники перераховуються автоматично.

12. Важливим є те, що список статей можна зберігати (експортувати) у різних форматах, а програма зберігає історію пошуків з усіма результатами.

Для визначення публікаційної активності кафедр слід виконати пошук статей кожного науковця кафедри (за прикладом рис. 1.5) та записати результати у таблицю. Приклад такої таблиці для 2-х кафедр ІПТДМ показано нижче.

Таблиця 1.1 – Результати пошуку за допомогою програми *Publish or Perish*

№	Кафедра	ПІБ	Число публікацій (взагалі)	Число публікацій за 2013–2014 рр.	Індекс Гірша	g-індекс
1	Інформаційних технологій проектування в машинобудуванні	Тонконогий В.М.	45	8	4	4
2		Колесникова Е.В.	23	4	2	3
3		Павлишко А.В.	6	0	0	0
4		Синько И.С.	2	0	0	0
5		Тигарев В.М.	6	0	1	1
6		Бовнегра Л.В.	14	1	1	1
7		Савельева Е.В.	2	2	0	0
8		Лебедев Б.В.	3	0	0	0
9		Якимов А.А.	3	1	1	1
10		Вайсман В.А	31	4	4	6
11		Рязанцев В.М.	11	1	2	2
12		Барчанова Ю.С.	2	2	0	0
13	Металорізальні верстати, метрологія та сертифікація	Оборский Г.А	19	3	1	2
14		Костенко В.Л.	10	0	2	2
15		Тихенко В.Н.	19	0	2	2
16		Моргун Б.А.	2	0	0	0
17		Слободяник П.Т.	8	0	2	5
18		Швагирев П.А	3	0	0	0
19		Чаругин Н.В.	1	0	0	0
20		Гнатюк А.П.	2	0	0	0
21		Гугнин В.П.	2	0	0	0
22		Луговская Е.А	2	0	0	0
23		Перпери Л.М.	3	0	0	0
24		Огиенко М.С.	1	0	0	0
25		Волков А.А.	3	0	0	0

Можна виконати також пошук статей для всієї кафедри у одному запиті. Для цього слід ввести всі прізвища, об’єднуючи їх логічним оператором OR, а після цього визначити публікаційну активність кафедр з використанням програми Google Академія. Оскільки цей програмний продукт широко застосовується науковцями, розглянемо спосіб розширення можливостей Google Академія для відображення результатів публікаційної активності кафедр. Для цього були зареєстровані 6 акаунтів на Веб-сайті Google відповідно до кількості кафедр Інституту промислових технологій дизайну та менеджменту ОНПУ (табл. 1.2).

Таблиця 1.2 – Перелік акаунтів кафедр ІПТДМ

№	Кафедра	Акаунт
1	Інформаційних технологій проектування в машинобудуванні (ІТПМ)	*kafedra.itpm@gmail.com*
2	Металорізальні верстати, метрологія та сертифікація (МСМС)	*kafedra.mcmc@gmail.com*
3	Технології та управління ливарними процесами (ТУЛП)	*kafedra.tulp@gmail.com*
4	Технології машинобудування (ТМ)	*kafedra.tex.mash@gmail.com*
5	Управління системами безпеки життєдіяльності (УСБЖД)	*kafedra.ysbjd@gmail.com*
6	Технології конструкційних матеріалів і матеріалознавства (ТКММ)	*kafedra.tkmm@gmail.com*

Провівши пошук інформації для кожного з діючих співробітників кожної з кафедр Інституту промислових технологій дизайну і менеджменту, отримаємо результати публікаційної активності кожної кафедри інституту (рис 1.6 – 1.11).

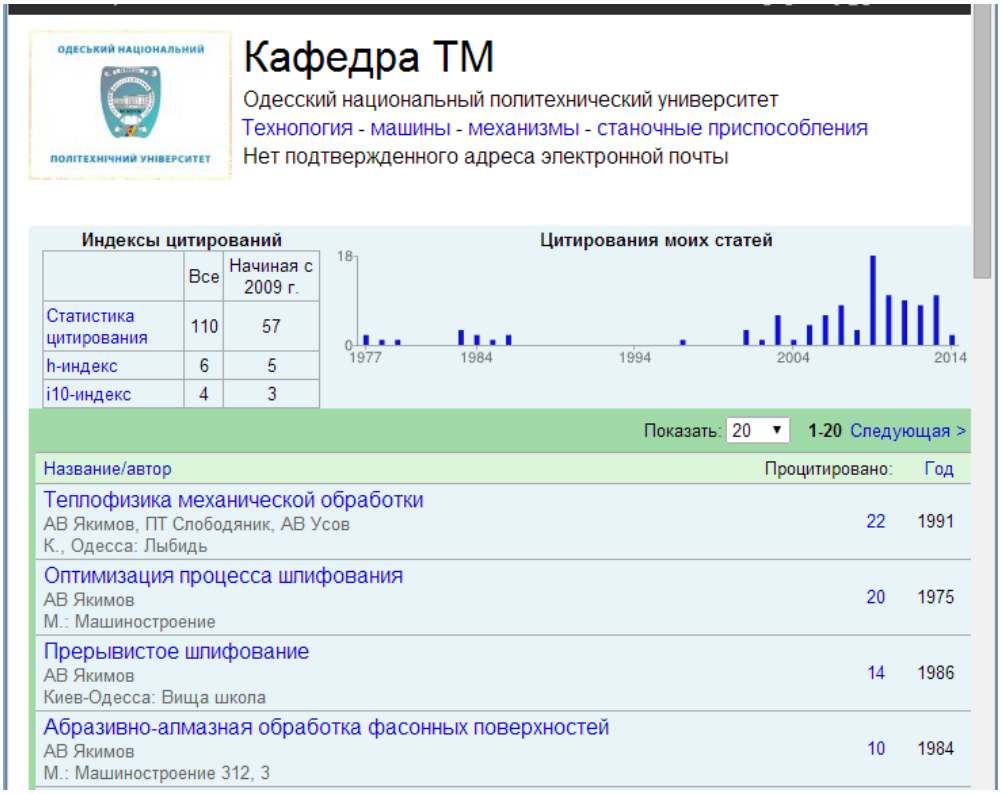


Рис. 1.6 – Результати пошуку по кафедрі ТМ – *kafedra.tex.mash@gmail.com*

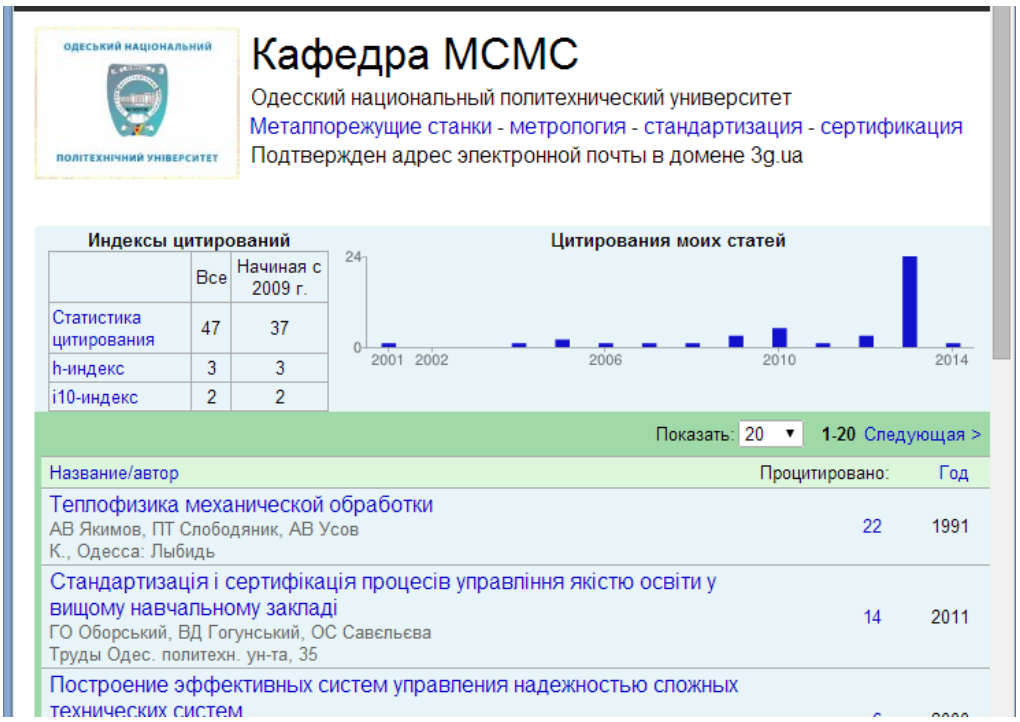


Рис. 1.7 – Результати пошуку по кафедрі МСМС – *kafedra.mcmc@gmail.com*

Рис. 1.8 – Результати пошуку по кафедрі ІТПМ – *kafedra.itpm@gmail.com*

Рис. 1.9 – Результати пошуку по кафедрі ТКММ – *kafedra.tkmm@gmail.com*

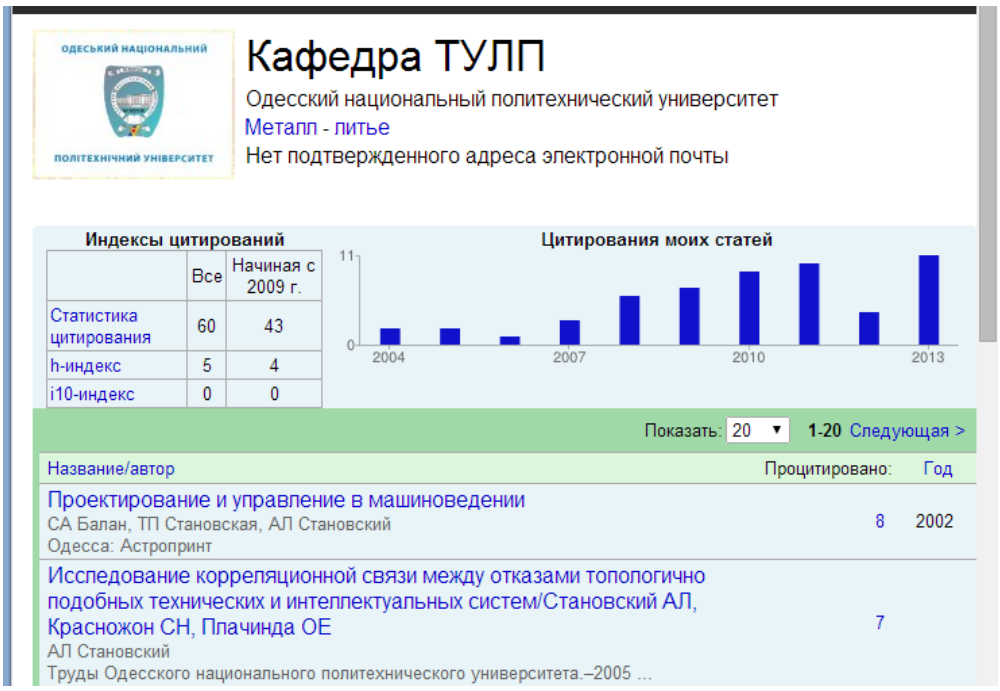


Рис. 1.10 – Результати пошуку по кафедрі ТУЛП – *kafedra.tulp@gmail.com*

Рис. 1.11 – Результати пошуку по кафедрі УСБЖД – *kafedra.ysbjd@gmail.com*

Порівняння публікаційної активності кафедр ІПТДМ у наукометричній базі даних Google Академія показує, що наукові результати у формі наукових статей науковців різних кафедр суттєво відрізняються за кількісними оцінками.

Представлений підхід до моніторингу публікаційної активності окремих науковців та їх колективів (лабораторій, кафедр, факультетів, тощо) дозволяє отримати досить об’єктивну інформацію щодо ефективності їх наукової діяльності, а також відкритості результатів досліджень для вітчизняної та міжнарод-

ної наукової спільноти. Отримані результати надають можливість оцінювання науковців та творчих колективів при вирішенні питань щодо розподілу грантів, присудження нагород та при рейтингуванні кафедр, факультетів, інститутів та університетів.

В ОНПУ за ініціативою проректора проф. Дмитришина Д.В. створені сторінки всіх кафедр і інститутів, що дозволяє відображати публікаційну активність науковців університету (рис. 1.12). Ці дані також відображені в проекті бібліотеки ім. В.І. Вернадського «Бібліометрика української науки».

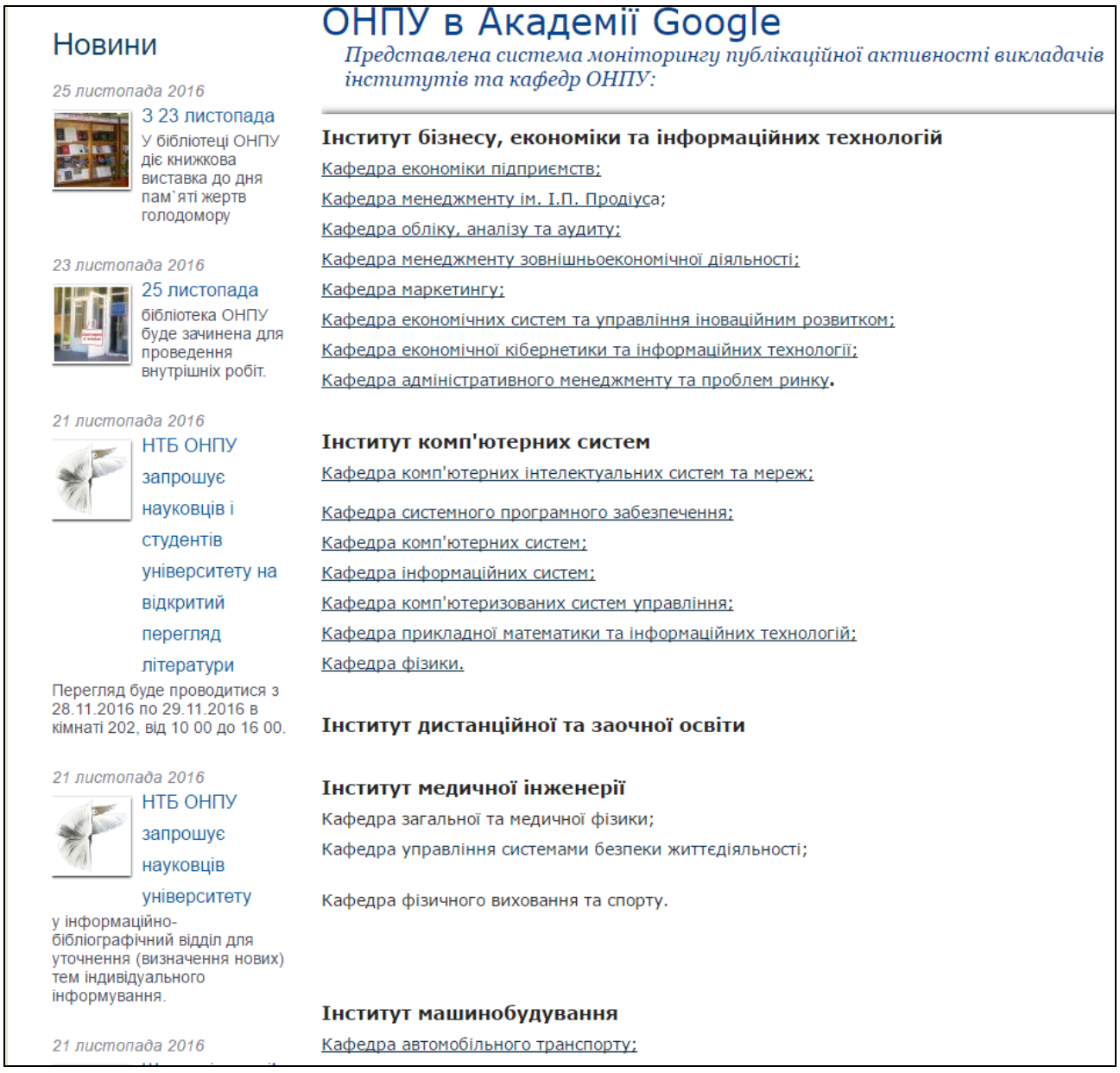


Рисунок 1.12 – Фрагмент Веб-сторінки «ОНПУ в Академії Google»

1.5 Наукометричні дослідження активністю публікацій як складова інноваційного розвитку університету

Теоретичні, функціональні і структурні зміни в різних областях знань транслюючись через наукові публікації у всесвітньому інтернет просторі, відображають тенденції розвитку наукових напрямків, нові отримані дані і досягнення конкретних дослідників [54]. Наявність доступної множини публікацій у світовій павутині створює умови для розвитку наукометричних досліджень щодо обгрунтування і застосування вимірювань в такий слабо структурованої області як наукові дослідження [55].

Дослідимо практичні аспекти роботи з некомерційними програмними продуктами "Publish or Perish" і "Google Академія" з розширенням їх області застосування для відображення результатів практики публікаційної активності викладачів кафедр. Для активізації практики публікаційної активності ВНЗ МОН України вводить низку заходів, спрямованих на інтеграцію в європейське і світове співтовариство університетів за рахунок подання у вигляді статей результатів досліджень вчених ВНЗ в зарубіжних журналах або в виданнях України, включених до зарубіжних наукометричних баз даних [11].

Серед найбільш вагомих заходів слід відмітити наступні. Змінено вимоги до видань при включенні до переліку фахових видань для створення умов відповідності цих видань міжнародним вимогам [49 – 53]. Конкурс щодо фінансування проектів наукових досліджень і розробок з 2013 проводиться з урахуванням числа публікацій, що індековані в Scopus і інших міжнародних наукометричних базах. Посилені вимоги «до планування дисертаційних досліджень, формулювання їх тематики, зокрема щодо формулювання теми, новизни, предмета і об'єкта дослідження». Суттєвою складовою дисертаційних досліджень є публікації у фахових журналах, а також в електронних виданнях [49]. Вимоги до публікації встановлюють: для докторської дисертації в цілому не менше 20 професійних публікацій, з них «не менше чотирьох публікацій у наукових періодичних виданнях інших держав з наукового напряму дисертації»; для кан-

дидатської дисертації – відповідно не менше 5 професійних публікацій і однієї статті у виданнях інших держав. До публікацій в наукових виданнях інших держав можуть прирівнюватися публікації у фахових виданнях України, включених до міжнародних наукометричних бази » [52].

До критеріїв оцінки діяльності ВНЗ включений показник «чисельність науково-педагогічних працівників, які мають публікації у виданнях іноземних держав або у виданнях України, включених до міжнародних наукометричних баз в звітному навчальному році» [51].

Зазначені вимоги до наукових публікацій результатів дисертаційних досліджень, безпосереднє оцінювання ВНЗ за кількістю публікацій, що входять в міжнародні наукометричних баз, а також формування нових державних вимог щодо акредитації, трансформують публікаційну активність вчених ВНЗ з особистої зацікавленості професорсько-викладацького складу в один з найважливіших показників діяльності вищих навчальних закладів . Це означає, що планування набору абітурієнтів, вибори викладачів, фінансування наукових досліджень будуть базуватися на даних про публікації та показниках цитування. Тому, очевидно, для управління цим проектом необхідно створити інформаційно-аналітичну систему моніторингу активністю публікацій вчених вузів України [10]. Саме ця діяльність повинна стати важливим кроком активізації виходу на міжнародний рівень: «Кожен вчений повинен знати число своїх публікацій і їх оцінку колегами у вигляді цитування. Узагальненим показником рівня цитування наукових публікацій є індекс Гірша »[11].

Актуальність дослідження активності публікацій ВНЗ пов'язана ще і з тим, що існують системи проведення рейтингу кращих університетів світу за версією ARWU (Академічний рейтинг університетів світу) Інституту вищої освіти Шанхайського університету Цзяо Тун [56] і версії QS World University Rankings [57] . Обидві системи складання рейтингу університетів обов'язково враховують якість і кількість публікацій. У Шанхайському рейтингу питома вага активністю публікацій університетів становить 60%; в рейтингу QS – 20%. На жаль, українські університети в цих рейтингах значно відстають від провідних

університетів США, Канади, Англії та Німеччини. Крім зазначених вище систем рейтингування існують і інші аідходи щодо встановлення рангів ВНЗ [58? 59].

Доступ до множини публікацій світової спільноти вчених формує нове ставлення до такої слабо структурованої області як бази даних наукових публікацій. Особливу увагу слід приділити якості публікацій – не тільки з точки зору новизни і практичної значущості досліджень, а й в плані оформлення та подання тексту статей на прийнятному англійською мовою.

Розглянемо основні параметри оцінки результативності публікацій.

Останнім часом найчастіше застосовуються: імпакт-фактор індекс цитування та *h*-індекс [39].

Індекс Гірша або *h*-індекс є кількісною характеристикою продуктивності одного учасника, групи вчених, університету або країни в цілому, що визначається на основі кількості публікацій і числа цитувань цих публікацій [12]. Для визначення індексу Гірша публікації ранжують у порядку за зменшенням числа посилань. Потім, визначають ту статтю, ранг якої збігається з числом її цитувань. Це число і є *h*-індексом, який, взагалі-то, не має фізичного тлумачення. Цей показник розділяє статті на дві частини. Статті в першій частині мають число цитувань, що перевищує ранг статті. Друга частина включає інші статті.

Індекс Гірша може обчислюватися з використанням як відкритих наукометричних баз в Інтернеті (наприклад, Google Scholar, Science Index (eLIBRARY.ru), ADS NASA), так і баз даних з платною підпискою (наприклад, Scopus або Web of Science) [41].

Індекс цитування має подвійне тлумачення [11]. В Україні це поняття визначає число цитувань публікацій. Сучасне тлумачення індексу цитування пов'язано з англомовної калькою цього поняття. Під індексом цитування розуміється реферативна база даних наукових публікацій, в якій виконується індексація посилань, зазначених в пристатейних списках публікацій і надається кількісна оцінка показників цих посилань (загальне число цитувань, індекс Гірша та ін).

Імпакт-фактор (ІФ або IF) – чисельний показник наукового рівня журналів [11]. З 1960-х років він щорічно розраховується Інститутом наукової інформації (англ. Institute for Scientific Information, ISI) і публікується в журналі «Journal Citation Report». Розрахунок імпакт-фактора заснований на трирічному періоді. Наприклад, імпакт-фактор журналу в 2015 році I_{2015} обчислюється за формулою:

$$I_{2015} = A / B,$$

де A – загальне число цитувань у 2015 році статей, опублікованих в даному журналі в 2013 і 2014 роках; B – число статей, що опубліковані в даному журналі в 2013–2014 роках.

На основі ІФ (в основному в інших країнах, але останнім часом і в Україні) оцінюють рівень журналів, якість статей, опублікованих в них, дають фінансову підтримку дослідникам і приймають співробітників на роботу. Імпакт-фактор має хоча і велике, але неоднозначне тлумачення впливу на оцінку результатів наукових досліджень.

Наукометричні бази і відносна достовірність даних.

Міжнародна практика наукометричних досліджень сьогодні найбільш часто базується на використанні двох баз даних: Web of Science і Scopus. Широко відомі також наукометричних баз даних: Springer, Begell House Inc., Pleiades Publishing, Kluwer і ін. Всі вони є комерційними базами.

Серед некомерційних наукометричних баз з технічних наук можна назвати наступні: Science Direct, Copernicus, Science Index, DOAJ, BASE, Driver, MLibrary, WorldCat, FreeFullPDF, arXiv, Google Serch і ін. [4].

Широко застосовується відома програма Publish or Perish, яка є пошуковою системою і дозволяє виконувати пошук публікацій на прізвище автора [38]. Результатом роботи системи є повний комплект наукометричних показників по публікаціям автора – від індексу Хірша до числа співавторів в знайдених статтях.

Переваги і недоліки

У табл. 1.3 на прикладі видання «Праці Одеського політехнічного університету», наведені ці дані з урахуванням того, що автори в своїх пристатейному списках літератури не завжди дотримуються загальноприйнятих правил написання назви видання. Множина публікацій в цьому виданні є слабо структурованою множиною даних. Тільки один атрибут – назва видання – має 16 різних значень. Якщо додати до цієї невизначеності ще й різні варіанти написання (переведення) прізвищ, оскільки вказане видання містить статті на трьох мовах (російською, українською та англійською), то невизначеність і варіабельність атрибутів статей може збільшитися на порядок.

Таблиця 1.3 – Показники цитування для видання «Праці Одеського політехнічного університету» в пошуковій системі Publish or Perish (дані 30.11.2016)

№	Варіанти назв видання в статтях	Статей	Цитат	h–індекс
1	Тр. Одес. политехн. ун-та	210	802	10
2	Труды Одесского политехнического университета	304	373	7
3	Праці Одеського політехнічного університету	390	91	4
4	Труды Одес. политехн. ун-та	20	52	4
5	Труды ОНПУ	11	17	3
6	Праці Одес. політехн. ун-ту	2	3	1
7	Пр. Одес. політехн. ун-ту	1	0	0
8	Пр. ОНПУ	1	2	1
9	Праці ОПУ	1	2	1
10	Праці Одес. держ. політехн. ун-ту	1	2	1
11	Труды ОГПУ	25	33	4
12	Тр. ОГПУ	7	15	2
13	Труды ОПУ	8	23	3
14	Тр. ОПУ	2	4	2
15	Odes'kyi politechnichnyi universytet. Pratsi	27	54	4
16	Сборник ОНПУ	1	0	0

Станом на 04.12.2016 р. пошукова система Google Scholar для видання «Праці Одеського політехнічного університету» відображає такі загальні наукометричні показники наукового видання: загальне число посилань – 2349; індекс Гірша – 17; i_{10} – індекс становить 36 (рис. 1.13).

Результати, наведені в табл. 1.3, показують, що слід звертатися до інших пошукових систем, в яких можна здійснювати пошук по багатьох атрибутах. При цьому слід зазначити, що пошук за прізвищем автора є найбільш достовірним.

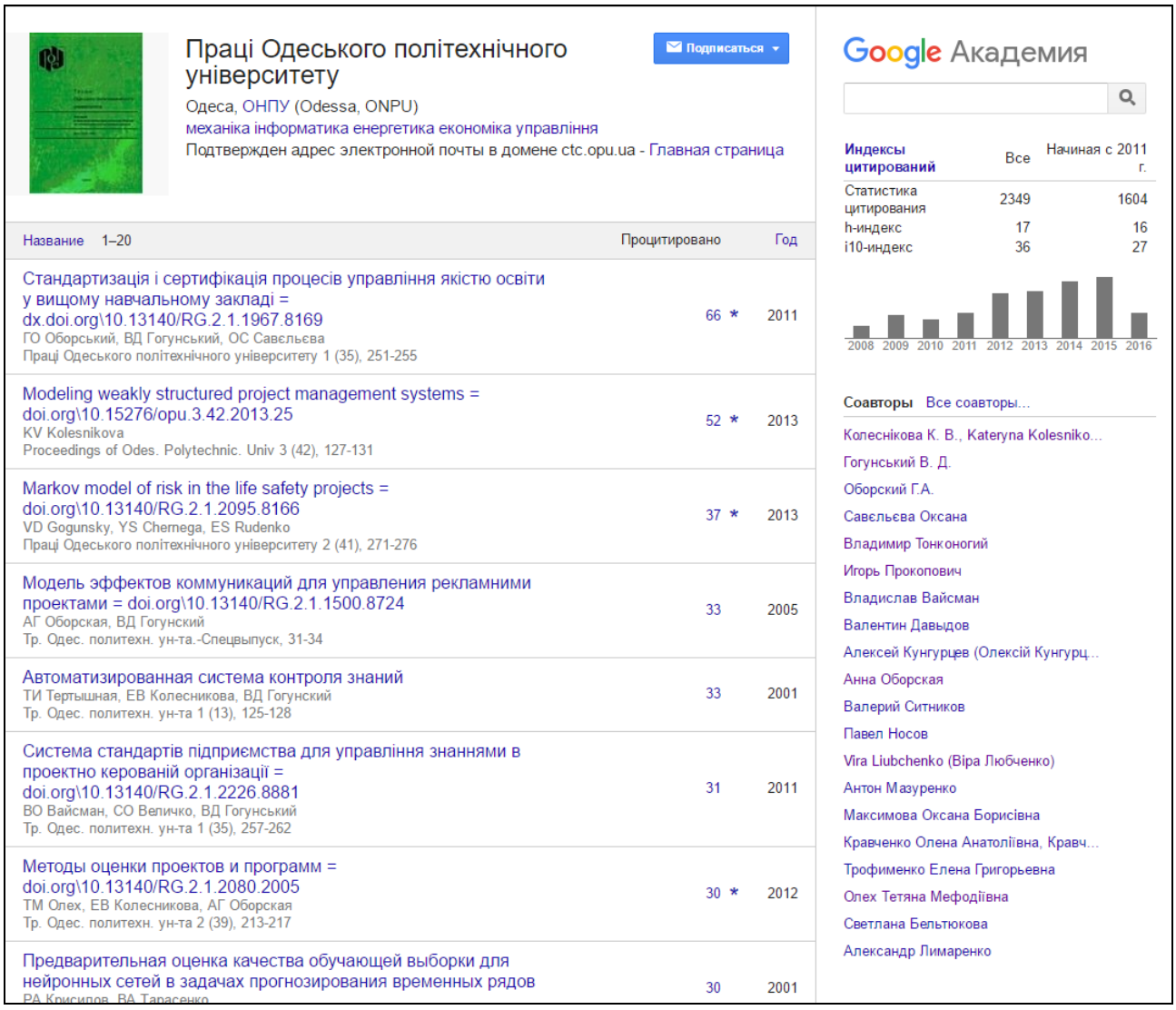


Рисунок 1.13 – Фрагмент відображення наукометричних даних видання «Праці Одеського політехнічного університету» у базі Google Scholar

Покажемо результати оцінки активністю публікацій на прикладі кафедр Інституту промислових технологій, дизайну та менеджменту (ІПТДМ) в базі публікацій Google Академія. Оскільки цей програмний продукт широко застосовується окремими вченими, розглянемо спосіб розширення можливостей

Google Академія для відображення результатів активністю публікацій викладачів кафедр.

Як вказано раніше, були зареєстровані 6 акаунтів на Веб-сайті Google, відповідно до числа кафедр ІПТДМ (табл. 1.2). В результаті пошуку даних для кожного із співробітників кафедр ІПТДМ отримали результати активності публікацій всіх кафедр інституту. На рис. 1.14 показана схема визначення Індексу Гірша для кафедр ІПТДМ.

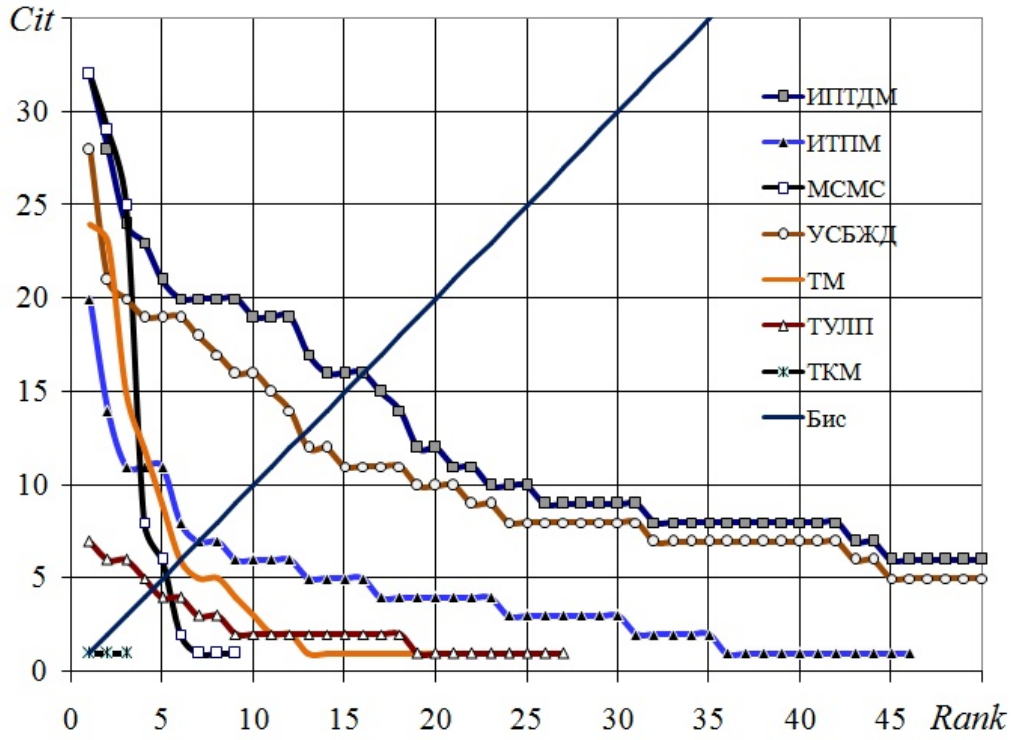


Рисунок 1.14 – Определение индекса Гирша для кафедр ИПТДМ

До переваг даного методу визначення результативності наукових публікацій слід віднести, перш за все, відсутність кореляції між числом публікацій і рівнем оцінки колег по науковому напрямку новизни і значущості публікації. Більш того, збільшення автором кількості статей з близької тематикою веде до зменшення індексу Гірша через розподіл цитувань між усіма статтями, які містять близькі за змістом дані. Хоча при цьому загальне число цитувань може навіть збільшитися.

Друга перевага оцінки наукової результативності підрозділів за допомогою Google Академія полягає в тому, що автоматично виключається проблема подвійного обліку публікацій, що має місце при простому підсумовуванні даних про публікації окремих співробітників кафедр. Це пояснюється тим, що основною інформаційною одиницею в базі даних є стаття, а не автор. Тому у разі наявності великої кількості співавторів на кафедрі йде в залік саме стаття. Просте підсумовування даних за публікаціями кафедр, як правило, не збігається з підсумковим результатом для факультету через наявність спільних статей співробітників різних кафедр.

Безсумнівним достоїнством Google Академія є можливість виконання розширених запитів із записом результатів пошуку в один загальний список публікацій. При цьому отримані дані можна коригувати і уточнювати в ручному режимі за допомогою інтуїтивно зрозумілого для користувача інтерфейсу.

Відкритість і прозорість даних дозволяють запрошувати своїх колег в свою бібліотеку. Це особливо корисно при роботі з аспірантами, магістрами, а також при спілкуванні з колегами з інших країн. Більш повна картина активності публікацій колективів кафедр представлена в табл. 1.4 (дані 1.09.2015). Зауважимо при цьому, що в даному порівнянні результативності та вагомості наукових публікацій не наведені дані про загальну кількість публікацій – немає сенсу аналізувати баласт (науковий шум). Слід оцінювати науковий результат, який виражається у визнанні наукової цінності наших публікацій колегами в формі цитування. Для формування загального уявлення про рівень активністю публікацій кафедр в табл. 1.4 показані також дані по інституту (ІПТДМ) і університету (ОНПУ).

При оцінці даних табл. 1.4 звертає на себе увагу зміна інтенсивнности цитувань за останні 5 років. Напевно, не слід вводити себе в оману про поліпшення нашої роботи. За період з 2010 року суттєво змінилися умови подання матеріалів публікацій в середовищі Інтернет: більшість наукових видань розробили свої Веб-портали, активно розширюються різні репозитарії, багато наукових видань України почали роботу по входженню до наукометричних баз даних, які

орієнтовані на платне або безкоштовне надання інформаційних послуг. Все це призводить до того, що присутність науковців України в Інтернет-просторі стає дедалі помітнішою.

Таблиця 1.4 – Загальні результати цитування статей (Google Академия)

№	Кафедра	Дані цитування		Індекс Гирша	
		Всі статті	З 2009 р.	Всі статті	з 2009 р
1	ІТПМ	201	185	7	7
2	МСМС	70	60	4	4
3	ТУЛП	62	44	4	4
4	ТМ	124	71	6	5
5	УСБЖД	565	528	12	12
6	ТКММ	3	2	1	1
7	Інститут ПТДМ	767	628	16	13
8	ОНПУ	3326	2157	20	15

Доступ до множини публікацій світової спільноти вчених формує нове ставлення до такої слабо структурованої області, як наукометричних баз даних публікацій. Навіть світові лідери в наданні наукометричних послуг, такі як Scopus, представляють дані у формі: «as is» (як є). Такий підхід не є продуктивним через відсутність зворотного зв'язку між авторами та командою супроводу наукометричних баз. Для підвищення достовірності визначення числа статей для університетів і організацій авторам публікацій слід надати можливість інтерактивного уточнення метаданих своїх статей.

1.6 Узагальнена схема формування рейтингів ВНЗ

Компетентнісний підхід в освіті переорієнтує освітянську парадигму з знань, умінь та навичок на ставлення і цінності у вигляді сформованих компетентностей, що є інновацією, яка спроможна розв’язати суперечності між запитами ринку праці та існуючою системою освіти [60]. Механізм взаємодії вищих навчальних закладів (ВНЗ) з зацікавленими сторонами реалізується за допомогою рейтингів ВНЗ, які формуються незалежними агенціями (рис. 1.15).

Промисловість, яка є Замовником і споживачем основного продукту ВНЗ – фахівців з вищою освітою, визначає зміну парадигми відбору претендентів на

робочі місця. Сьогодні це сукупність цінностей, технічних навичок, поведінкових елементів до претендентів на участь у сучасних формах організації виробничих процесів, наявність комунікативних властивостей особистості, які б задовольняли потребам динамічного середовища виконуваних проектів [61 – 63].

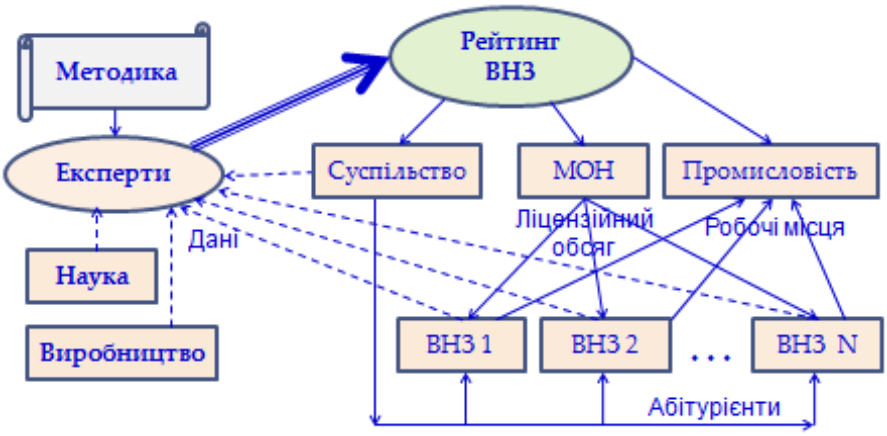


Рисунок 1.15 – Узагальнена схема формування рейтингів ВНЗ

Загальновизнаним критерієм ефективності роботи ВНЗ, крім якості підготовки фахівців – академічною репутацією (40 %) та оцінкою роботодавців (10 %), є обсяг і рівень науково-дослідницької діяльності, який безпосередньо відображається рівнем публікаційної активності викладачів [57]. Вагомість складової цитування публікацій у рейтингу QS складає 20 % (рис. 1.16).

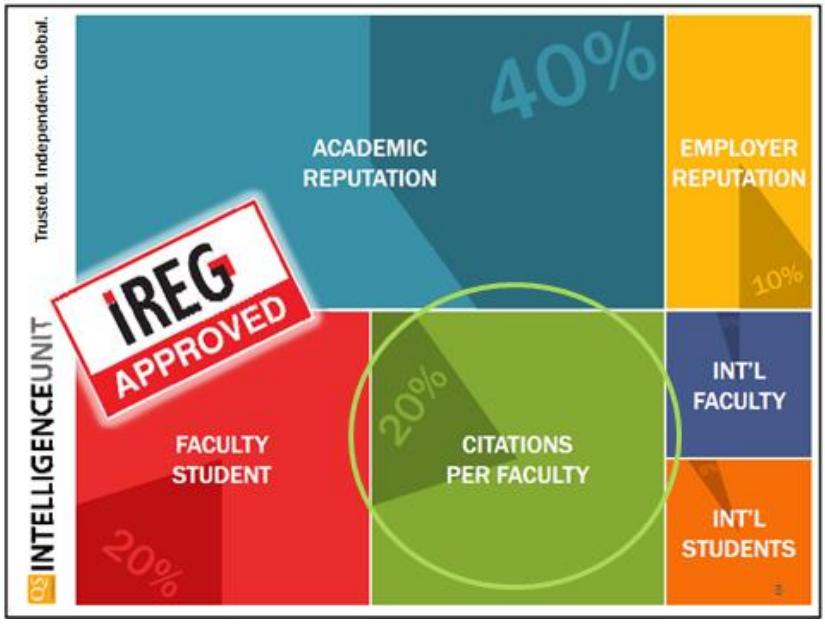


Рисунок 1.16 – Вагомість складових рейтингу ВНЗ за версією QS

Використання зазначених вище наукометричних БД дозволяє здійснити пошук повного комплекту наукометричних показників за публікаціями авторів – від індекса Гірша до числа співавторів у знайдених статтях [43]. Попередня оцінка публікаційної активності може бути спрямована на реалізацію проектів управління активізацією публікаційної активності науковців ВНЗ.

1.7 Узагальнена схема наукометричних баз у світовій Web-мережі

Процеси глобалізації, інтеграції наукових досліджень, становлення інформаційних технологій щодо організації міжнародних наукометричних баз даних та електронних бібліотек з доступом до наукових публікацій породжують нові можливості і завдання в сфері освітньої та наукової діяльності у вищій школі України. Одним з напрямів цієї діяльності є визначення узагальненої оцінки якості та результатів наукових досліджень окремого вченого, кафедри, факультету, університету і вищих навчальних закладів України в цілому. Можливість аналізу публікацій у світовій павутині створює умови для розвитку наукометричних досліджень – наукового напрямку з оцінки та застосування вимірювань у такий слабо структурованої галузі як наукові дослідження [6].

Зростання вимог до теоретичного і практичного значенню наукових досліджень обумовлює необхідність ефективного використання сучасних інформаційних технологій та методів проведення наукового пошуку опублікованих результатів досліджень [7]. Теоретичні, функціональні та структурні зміни в різних областях знань певним чином відображаються у наукових публікаціях. Саме сукупність публікацій є основою для формування нових знань. Світовий досвід взаємодії спільноти вчених з інформаційним середовищем всесвітньою Web-павутини свідчить про доцільність застосування деяких показників продуктивності наукової діяльності.

Сьогодні рівень наукоємності і досконалості систем різного призначення визначені у світі як ключовий фактор формування конкурентоспроможності держави та бізнесу [64 – 70]. Тому актуальним завданням є публікація резуль-

татів досліджень у провідних фахових зарубіжних журналах або у вітчизняних виданнях, які включені в міжнародні наукометричні бази [71].

Розглянемо особливості застосування наукометричних баз:

– аналіз характеристик та основних властивостей наукометричних баз та індикаторів цитування наукових публікацій;

– виявлення особливостей відображення наукових статей у наукометричних базах даних (БД);

– визначення найбільш застосовуваних характеристик продуктивності наукової діяльності у світовому науковому співтоваристві;

– рекомендації щодо ознайомлення широкого кола науковців з показниками оцінки значущості наукових публікацій.

Тенденції економіки сучасного інформаційного суспільства такі, що рушійною силою інноваційного розвитку суспільства стає наука [66]. Поширення і просування наукових досягнень здійснюється через інформаційні канали, серед яких, як найбільш значущі, можна виділити засоби масової інформації, Інтернет, мобільні технології (рис. 1.17).

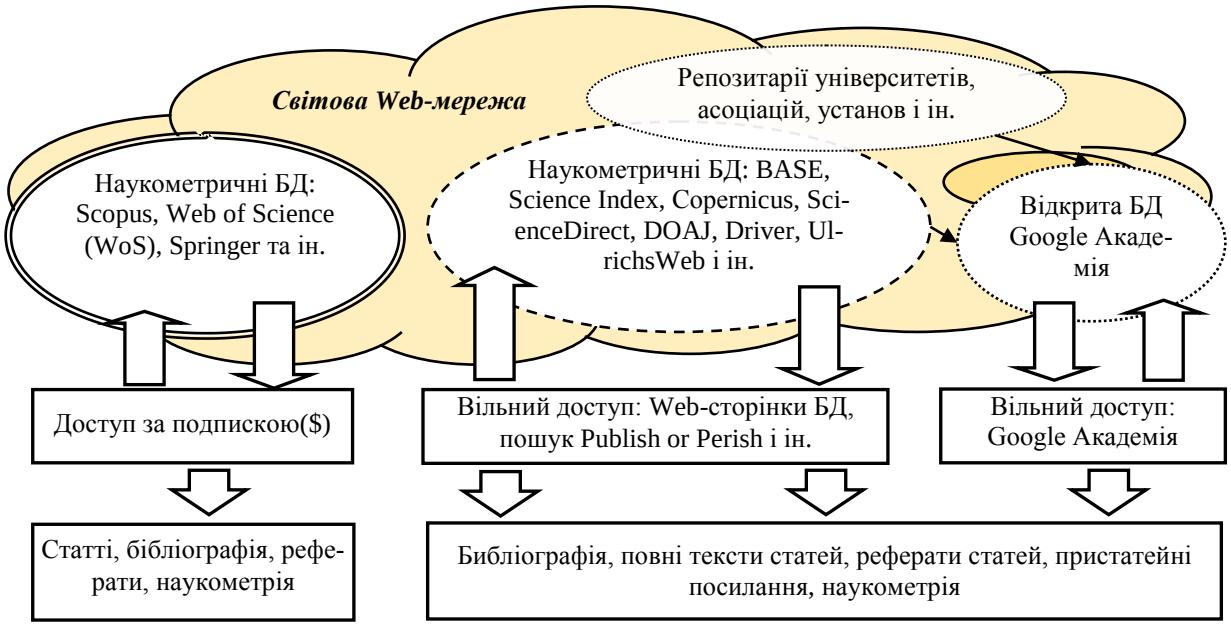


Рисунок 1.17 – Узагальнена схема світової Web-мережі

Наукометричні БД є основними осередками трансформації знань і каналами подальшого застосування наукових результатів, як головної інформаційної та соціальної характеристики країни, університету, наукового колективу або окремого науковця. Сьогодні рівень наукоємності та досконалості систем різного призначення визначено у світі як ключовий механізм формування конкурентоспроможності держави та бізнесу.

1.8 Висновки до розділу 1

Наукометричні бази даних є спеціалізованими засобами автоматизації наукометричної діяльності. Їхня поява стала наслідком експоненціального зростання науки в середині XX століття, коли наукові співробітники змушені були витрачати майже половину свого робочого часу на інформаційну діяльність. До недавнього часу про наукометричні базах знали і використовували їх в основному бібліотечні працівники різних країн. Сьогодні ж, як мінімум, кожен науковий співробітник України знає про них.

Найбільш відомими і великими наукометричними базами даних є Scopus та Web of Science. Серед некомерційних наукометричних баз даних, в яких індексуються наукові публікації можна назвати наступні: Copernicus, BASE, DOAJ, Science Index, WorldCat, MLibrary.

Активність публікації наукових співробітників є одним з основних факторів, який враховується при визначенні світових рейтингів вищих навчальних закладів (ВНЗ). Тому МОН України цілеспрямовано орієнтує публікаційну діяльність наукових співробітників на входження до світового наукового співтовариства. Кілька важливих кроків було розроблено в цьому напрямку: змінені вимоги до наукових видань та до фінансування проектів наукових досліджень і розробок; посилено вимоги до планування дисертаційних досліджень, формулювання їх тематик; змінені критерії оцінки ВНЗ України та інші. І заключним етапом є створення інструментів «вимірювання» активністю публікацій вчених з подальшим формуванням інформаційно-аналітичної системи моніторингу ак-

тивністю публікацій вчених ВНЗ України. З цим етапом і пов'язана дана робота по вилученню метаданих публікацій з наукометричних баз даних.

Автоматизація вилучення метаданих публікацій з різних НМБД дозволить виконувати моніторинг активності публікацій наукових співробітників.

2 СТРУКТУРА ТЕХНОЛОГІЧНОГО КОМПОНЕНТА МЕТОДОЛОГІЇ ПРОЕКТНО-ВЕКТОРНОГО УПРАВЛІННЯ.

2.1 Веб-інтерфейс як основний доступ до інформації з наукометричних баз

Інтерфейс включає сукупність можливих способів і методів взаємодії двох систем, пристроїв або програм для обміну інформацією між ними з певними їхніми характеристиками, а також характеристиками сполук, сигналів обміну і ін. У разі, якщо одна із взаємодіючих систем – людина, частіше говорять лише про другу систему, тобто про інтерфейс тієї системи, з якою людина взаємодіє (інтерфейс, що призначений для користувача). Одним із прикладів призначеного для користувача інтерфейсу є Веб інтерфейс програм у всесвітній мережі Інтернет. Веб-інтерфейс – це сукупність засобів, за допомогою яких користувач взаємодіє з Веб додатком.

На сьогодні налічується значна кількість міжнародних наукометричних баз даних, які розрізняються структурою і способом зберігання інформації. Програмний інтерфейс для доступу до кожної базі, якщо і існує, то часто не афішується. Не існує єдиного, універсального інтерфейсу, який підходив би до всіх баз. Але є один інтерфейс, який мають багато наукометричних баз даних і орієнтований він більше на користувача, ніж на програмне забезпечення.

Доступ до вмісту (в обмеженому вигляді) надає Веб-інтерфейс. Користувач, за допомогою Веб-браузера, завантажує Веб-сторінку певної наукометричної бази даних і, використовуючи пошук по заданих параметрах, отримує необхідну інформацію на сторінці.

В даному дослідженні пропонується витягувати програмним способом інформацію, орієнтовану на користувача (людини). Таким чином, імітується робота користувача, який завантажив би тисячі Веб-сторінок і зібрав би «вручну» інформацію певної структури до місцевої точки зберігання.

Класичним і найбільш популярним методом створення Веб-інтерфейсів є використання HTML із застосуванням CSS і JavaScript'у. Існує декілька технік для аналізу і обробки вмісту Веб сторінки:

– робота з Веб сторінкою, як зі звичайним текстом – вилучення вихідного коду сторінки і застосування утиліт для роботи з текстовою інформацією (наприклад, регулярні вирази);

– використання мов запитів для слабоструктурованих даних; Веб сторінки представлені мовою розмітки, що складається з іменованих тегів; прив'язуючись до цих тегів можна виконувати аналіз і обробку даних (XPath – мова запитів до xml подібних документів).

– побудова і робота з об'єктною моделлю документа (DOM) – не залежить від платформи і мови програмного інтерфейсу, що дозволяє отримати доступ до вмісту Веб документів, а також їх змінювати, аналізувати структуру і оформлення; даний спосіб імітує роботу Веб браузера.

Розвиток інтернет-технологій в області організації сховищ даних, сховищ і електронних бібліотек з наданням доступу до баз даних наукових публікацій, створює умови для розвитку досліджень в різних областях знань, які в певній мірі відображаються в наукових публікаціях. Саме множина публікацій становить основу формування нових знань. Розробка інтелектуального інтерфейсу для взаємодії з різними наукометричними базами даних дозволить істотно спростити пошук інформації [42].

2.2 Модель вилучення інформації з Веб сторінок

Застосування методу Веб скрапінга породжує задачу аналізу і ідентифікації слабоструктурованих даних. Слабоструктуровані представлення даних відрізняються відсутністю строгих структур таблиць і відносин в моделях реляційних баз даних, проте, ця форма даних містить теги та інші маркери для відділення семантичних елементів, а також для забезпечення ієрархічної структури

записів і полів в наборах даних [72]. Проблема полягає в аналізі інформації, яка міститься на Веб-сторінці.

Глобальна мережа Інтернет є найбільшим джерелом даних, велика частина яких представляється у вигляді Веб-сторінок, які не мають строго формалізованої структури. Витяганням інформації з таких джерел займаються такі великі корпорації як Google і Microsoft. Для якісного пошуку використовуються складні математичні моделі, семантичний аналіз та інші методи аналізу інформації. Тому дані, які надходять на вхід цих систем, повинні бути структуровані певним чином. Однак більшість наукометричних баз даних представлені у формі унікальних структур, що ускладнює отримання структурованих даних з подібних слабоструктурованих Веб сторінок.

Витяг структурованих даних з Веб сторінок зводиться до вирішення наступних завдань [73 – 76]:

– пошуку та отримання цільових сторінок для отримання інформації (проблема навігації);

– розпізнавання ділянок, що містять потрібні дані (проблема розпізнавання даних);

– пошуку структури знайдених даних (проблема пошуку загальної структури даних);

– забезпечення однорідності видобутих даних (проблема зіставлення атрибутів видобутих даних);

– об'єднання даних з різних джерел (проблема об'єднання даних).

Для вирішення завдання отримання даних на прикладі наукометричних баз даних [77], пропонується модель програмного забезпечення (рис. 2.1), яка складається з наступних компонентів:

– програма для отримання даних з конкретних НМДБ;

– блок фільтрів витягнутих результатів;

– база даних для кінцевих результатів.

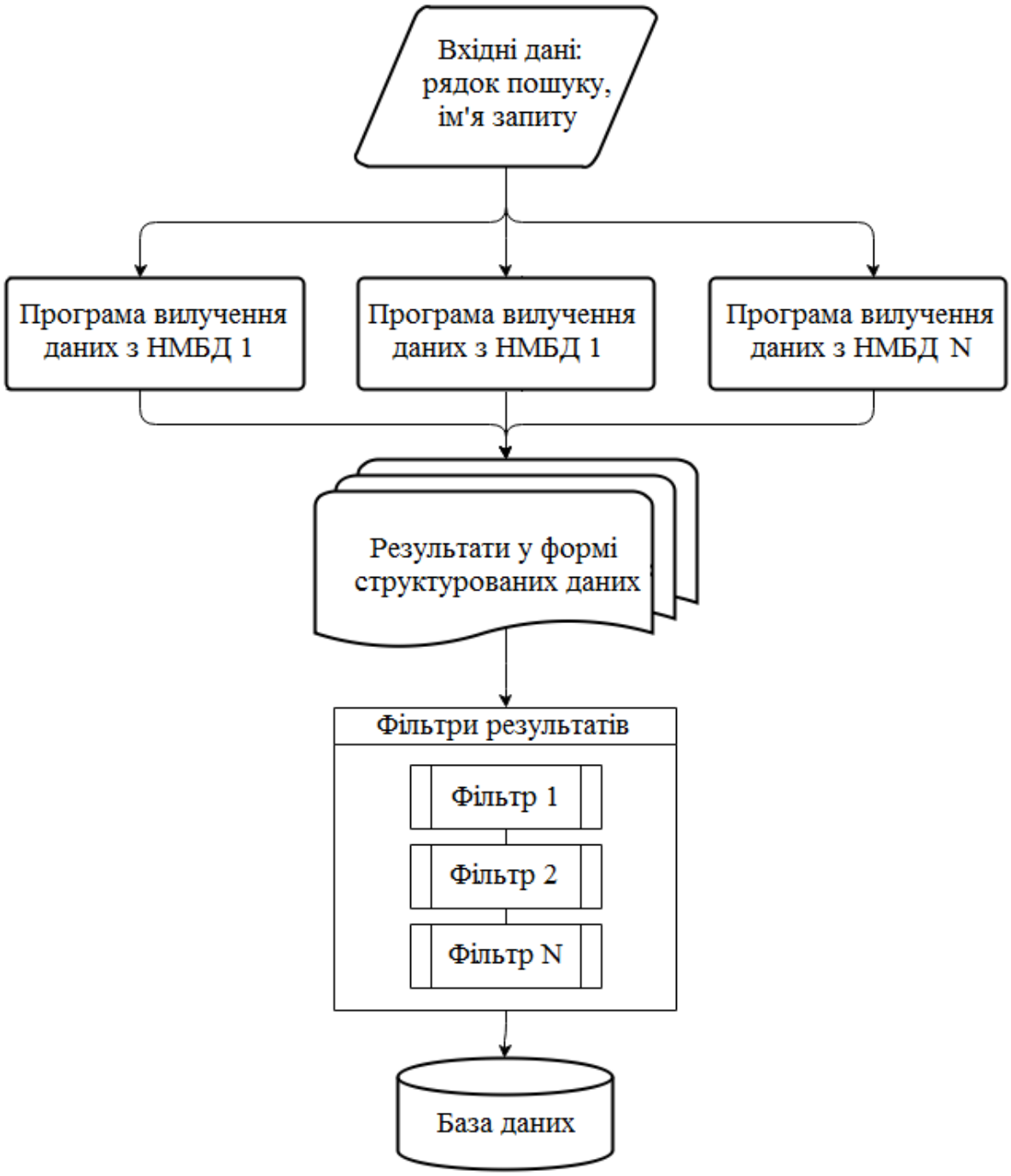


Рисунок 2.1 – Схема роботи програмного комплексу витягання інформації з наукометричних баз даних

Для кожної НМБД створюється окрема програма вилучення даних, оскільки всі бази мають різний інтерфейс і структуру. Ці програми містять в собі логіку роботи з конкретною НМБД, а також необхідні параметри, константні дані для виконання цієї роботи.

Після завершення роботи програм вилучення даних вихідні результати кожної з них збираються в загальний масив, який далі передається в блок фільтрів. Блок складається з одного або декількох фільтрів, які відкидають нерелевантні результати, згідно з деякими параметрами, специфічним для цього фільтра. Наприклад, результати програм вилучення даних можуть містити записи, що не відповідають запиту пошуку. Для цього можна використовувати фільтр, який залишатиме, тільки результати відповідні пошуковому рядку. Також можна використовувати фільтр для відкидання результатів однофамільців, видалення дублікатів і ін.

Після обробки блоком фільтрів набір результатів, що залишився записується в базу даних для подальшого представлення та аналізу.

Розглянемо докладніше роботу програми вилучення даних. На рис. 2.2 показана загальна схема її роботи. Кожна програма може мати відмінності в деталях через слабоструктурованих даних.

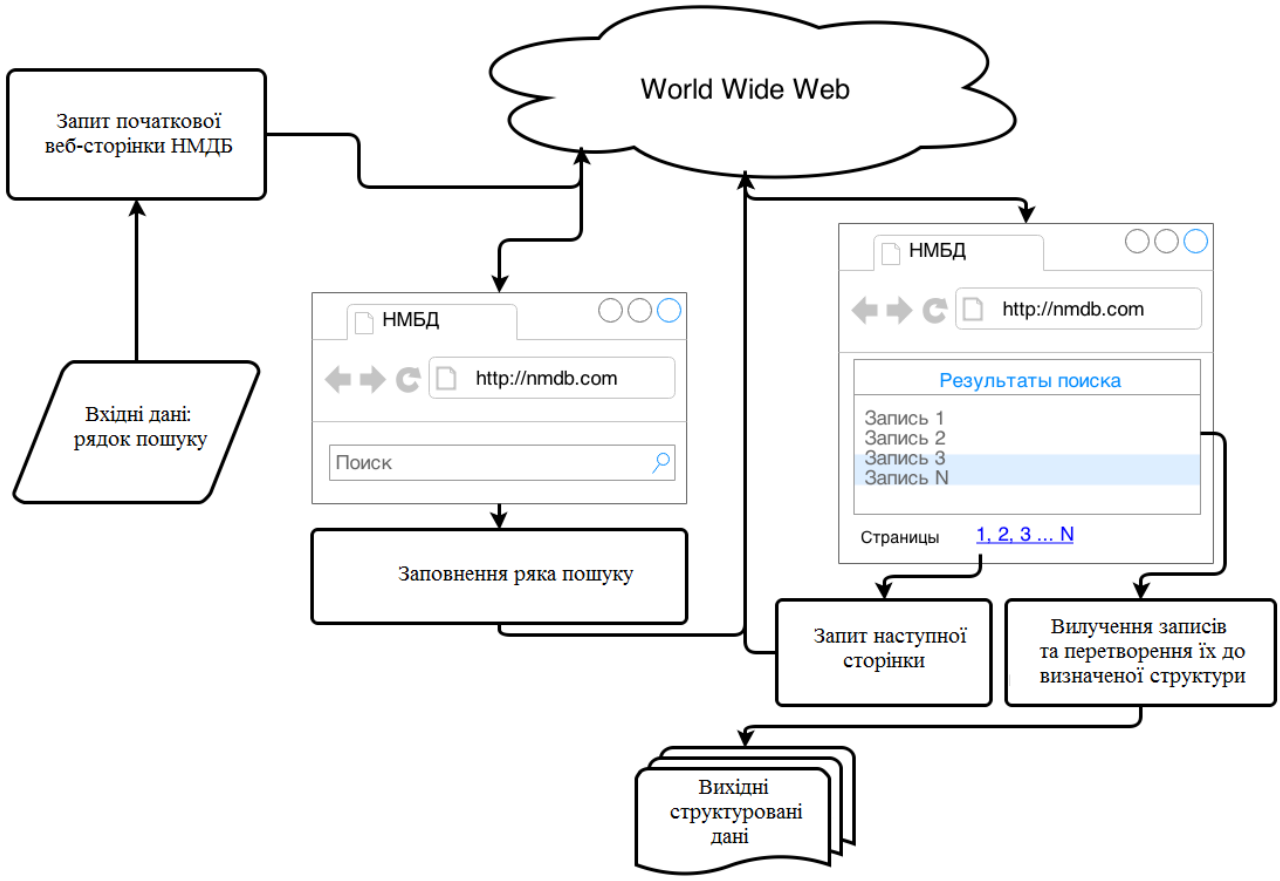


Рисунок 2.2 – Загальна схема роботи програми вилучення даних з НМДБ

На вхід програми подається рядок пошуку. Далі виконується запит на завантаження початкової сторінки конкретної НМБД, де програмним шляхом імітується робота користувача браузера – вводиться рядок пошуку в поле пошуку і виконується запит на видачу результатів. З Веб-сторінки результатів, за допомогою підпрограми, витягуються всі поля кожного запису і перетворюються в структуру строго певну програмним комплексом. Відсутні поля залишаються порожніми. Таким же чином витягуються посилання на інші сторінки, так як велика кількість результатів може бути розділене на сторінки. Далі виконуються запити на ці посилання, і процес повторюється спочатку, поки не будуть оброблені всі результати або спрацює обмеження на кількість, яке для кожної НМБД встановлюється окремо (один з параметрів). В кінці роботи програми вилучення даних отримуємо набір структурованих даних, готових для подальшої обробки.

Розглянемо сам процес добування інформації зі сторінки НМБД. Веб сторінка, яку повертає сервер, відформатована з використанням мови розмітки (в основному HTML), для подальшого відображення в тому чи іншому вигляді за допомогою спеціальної програми (Веб-браузер) (рис. 2.3).

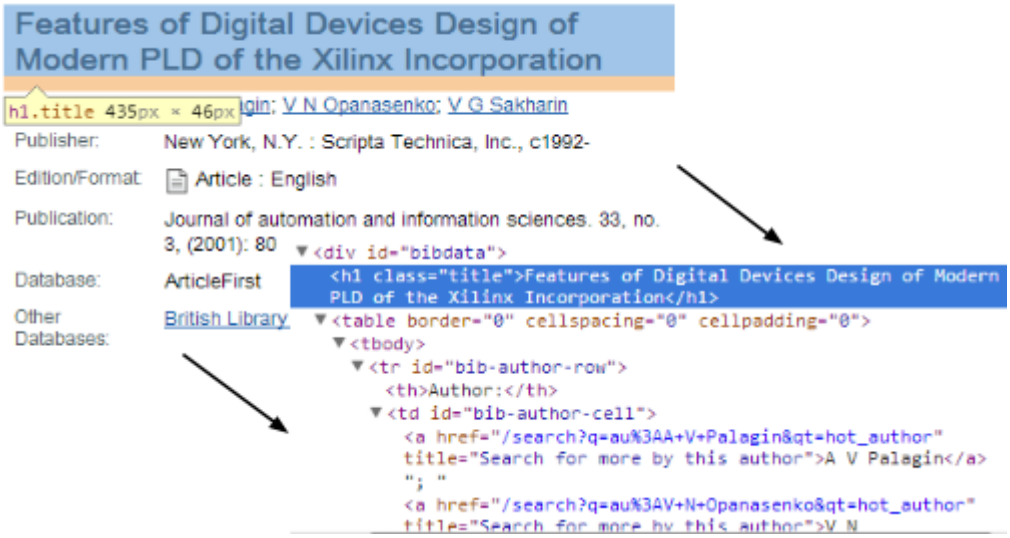


Рисунок 2.3 – Дані з Веб-сторінки і їх вихідний код мови розмітки HTML

На рис. 2.3 показаний приклад візуалізації Веб-браузером деякої області даних і вихідний код цих даних. Тут, наприклад, назва статті "Features of Digital Devices Design of Modern PLD of the Xilinx Incorporation" укладено в наступні спеціальні послідовності символів, звані тегами: *<h1 class = "Title"> Тут назва статті </ h1>*. Для вилучення цієї інформації, виконується пошук цих тегів і витягується їх вміст. Таким чином, заповнюється одне з полів результатів. Для автоматизації цього процесу, програми вилучення даних використовують мову запитів до елементів мови розмітки (Xpath).

Досліджуємо особливості вилучення даних з наукометричних баз даних, підтримуваних розробленим програмним забезпеченням. На поточний момент визначена наступна структура даних для кожної публікації (табл. 2.1).

Таблиця 2.1 – Структура вилученої інформацфї

Поле	*Опис*
Наукометрична база	Назва бази джерела публікації
Автори	Автори публікації
Назва	Назва публікації
Дата	Дата публікації
Джерело	Джерело публікації або видавництво
Опис	Анотація або короткий опис публікації
URL	Веб-посилання на публікації

BASE (*base-search.net*)

Наукометрична база даних BASE дозволяє виконувати пошук на різних мовах і не задає строгих правил щодо завдання пошукової послідовності (наприклад, ініціали автора можуть бути з точками або без них, а також разом).

Результати пошуку подаються в візуально структурованому вигляді, є наступні поля:

- назва публікації,
- автор (и),
- предмет,
- видавництво,

- рік видання публікації
- URL джерела публікації.

Але вихідний код на мові розмітки HTML має складну структуру і до того ж імена тегів залежать від мови інтерфейсу сайту. Тому перед початком роботи з цією базою, слід встановити мову інтерфейсу – англійська. Результати пошуку знаходяться всередині тегів-контейнерів з ім'ям класу "ResultsContent". Для кожного результату можна аналізувати його вміст: теги з ім'ям класу "ItemLeft_en" містять ім'я поля, а теги з ім'ям класу "ItemRight_en" – значення. Далі можна адаптувати цю інформацію під структуру даних (табл. 1) і отримуємо витягнуту запис.

Scopus (*scopus.com*)

Пошук у наукометричній базі даних Scopus виконується тільки на латиниці. При цьому для прізвищ та ініціалів є два різних поля введення. Ініціали слід вказувати з точкою. Робота з цією базою даних має особливості, в основному, через те, що результати пошуку – це інформація про автора. Тому для Scopus остаточна структуру даних раціонально розширити до 2 полів: кількість документів і *h*-індекс.

Результати пошуку видаються у вигляді таблиці з наступними полями:

- автор (и),
- кількість документів,
- предмет і ін.

Якщо автор має посилання на розширену інформацію, слід перейти за цим посиланням та записати цю інформацію у відповідне поле (URL). На сторінці розширеної інформації дані представлені у вигляді таблиці, що складається з трьох колонок: ім'я поля, роздільник, значення поля. Перебираючи рядки таблиці, можна заповнити вихідну структуру даних.

Science Index (*elibrary.ru*)

Пошук підтримується на багатьох мовах. Для більш ефективного пошуку по автору, використовується розширений пошук, де вказується прізвище автора та ініціали, які розділені прогалиною.

Результати пошуку подаються в таблиці, кожен рядок якої містить неструктуровану інформацію:

– назва статті,
– автор (и),
– джерело,
– URL; і
– дата публікації.

Для адаптації цієї інформації під загальну структуру слід застосувати наступні маніпуляції з рядком результату (рис. 2.4):

– назва і URL публікації витягуються із тега <b>, який знаходиться всередині тега <a>;
– автори публікації витягуються з тега <i>, який знаходиться всередині першого тега <font>;
– з другого тега <font> витягується дата і джерело публікації.

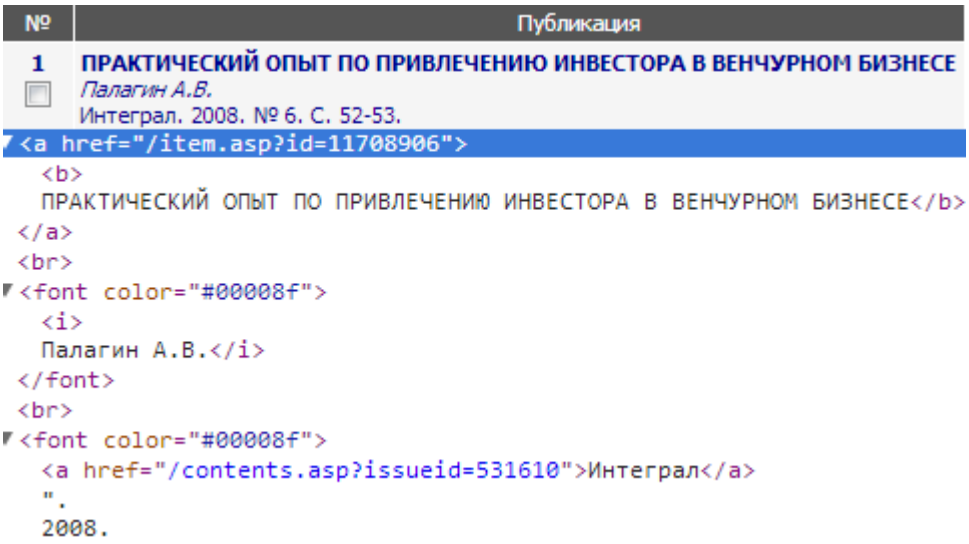


Рисунок 2.4 – Приклад рядка результату пошуку в наукометричної бази даних Science Index

Mlibrary (*lib.umich.edu*)

Наукометрична база даних Мічиганського університету Mlibrary надає пошук на латиниці і має розширений режим пошуку для завдання атрибутів, які є визначальними для пошуку. Доцільно використовувати параметр "Автор" для пошуку. Помічено, що запис ініціалів через пробіл видає більше результатів.

Результати пошуку видаються у вигляді списку з назвою публікації і посиланням на повний опис. Слід переходити по цих посиланнях і витягати інформацію з вмісту. Структура вихідної інформації представлена в наступному вигляді: теги з ім'ям класу "article-field-label" містять ім'я поля, теги з ім'ям класу "article-field-value" - значення. Проходом по всім полям можно вилучити інформацію, яка є необхідною для формування результату запиту(табл. 2.1).

WorldCat (*worldcat.org*)

Пошук по базі WorlCat також виконується на латиниці з використанням розширеного режиму, де вказуємо параметр "Автор" і "Формат публікації – стаття". Як і з базою Mlibrary, ініціали автора в рядку пошуку слід вказувати розділені пропуском.

Результати пошуку – список публікацій з коротким описом і посиланням на повний опис. Знову слід перейти по всіх посиланнях і працювати з інформацією на цих сторінках. Вміст сторінок цієї наукометричної бази даних має добре виражену структуру, що є дуже рідким для Веб-сторінок. Тут кожне поле має свій ідентифікатор, за яким можна вилучити певне значення. Наприклад, ідентифікатор "bib-author-cell" містить значення поля "Автори", а "bib-publisher-cell" – значення поля "Видавництво". Таким чином, можна легко заповнити свою локальну структуру даних (табл. 2.1).

2.3 Модель Веб скрапінгу для автоматизації вилучення даних

Множина даних в слабоструктурованої системі всесвітньої павутини утворює складну структуру організації інформаційних взаємодій, що змінюються в часі. При цьому деякі видання можуть бути включені в одну і більше наукометричних баз. Число публікацій постійно збільшується. Формати подання бібліографічних даних і в публікаціях, і в наукометричних базах суттєво відрізняються. Пошук публікацій в такому різнорідному неформалізованому середовищі часто доводиться робити тільки в «ручному» режимі.

Процес пошуку в даному середовищі є більш мистецтвом, ніж інформаційною технологією і залежить від умінь і навичок користувача. Проблема полягає в тому, щоб максимально формалізувати й автоматизувати цей процес.

Для розв’язання цієї проблеми потрібен спосіб отримання даних з наукометричних баз в структурованому вигляді для можливої подальшої їх обробки.

Дослідженнями в напрямку вилучення інформації з глобальної мережі Інтернет займаються великі компанії Google, Yandex, Microsoft. Вони використовують результати досліджень в реалізації пошукових машин, які є головним компонентом пошукових систем. Пошукова машина являє собою комплекс програм, призначений для пошуку інформації. Однією з головних функцій пошукових машин є отримання інформації з мережі. Далі відбувається обробка результатів, їх індексація для прискорення видачі результатів пошуку і підвищення його релевантності.

Основними компонентами підсистеми збору та вилучення інформації є:

– «Павук» (Spider) – програма для завантаження Веб-сторінок;
– «Краулер» (Crawler) – програма для автоматичного проходу по всіх посиланнях, знайденим на сторінці.

Павук викачує Веб-сторінки тим же способом, що і Веб-браузер, тобто імітується дія користувача. Але Веб-браузер відображає цю інформацію в графічному вигляді, а павук зберігає її для подальшої обробки. Краулер виділяє всі посилання, присутні на сторінці і переходить по всіх або по певних посиланнях, виходячи з заданих наперед умов пошуку. Слідуючи по знайдених посиланнях, він перенаправляє сторінки павуку для їх завантаження.

Робот Googlebot – це розроблена Google програма сканування Інтернету («павук»). Сканування є процесом, в ході якого робот Googlebot виявляє нові та оновлені сторінки для додавання в індекс. Google використовує величезну мережу комп'ютерів, щоб витягти вміст мільярдів Веб-сторінок. Робот Googlebot функціонує автономно і застосовує алгоритмічний процес: комп'ютерні програми визначають сайти, які потрібно сканувати, а також частоту сканування і кількість видобутих сторінок на кожному сайті.

Процедура сканування починається з отримання списку URL Веб-сторінок, який створюється на основі результатів попередніх сеансів сканування. Його доповнюють дані з файлів Sitemap, наданих Веб-майстром. Час відвідування таких сайтів робот Googlebot знаходить на кожній сторінці посилання і додає їх до списку сторінок, які потрібно сканувати. Всі нові й оновлені сайти, а також непрацюючі посилання позначаються для поновлення в індексі [78].

Основна мета – розробити спосіб отримання даних про публікації по параметру “Автор” з найбільш відомих наукометричних баз даних з можливістю розширення підтримуваних джерел [79]. Другорядним завданням є знайомство з найбільш відомими наукометричними базами даних.

Для автоматичного пошуку і вилучення даних використовується підхід, заснований на вживаному в пошукових машинах – Веб-скрапінг [80].

Веб-скрапінг – це процес добування інформації з Веб-сторінок, який фокусується на перетворенні неструктурованих даних в мережі (наприклад, у форматі HTML) в структурований формат даних, який може бути проаналізований і збережений. Веб-скрапінг також відноситься до автоматизації роботи у всесвітній павутині.

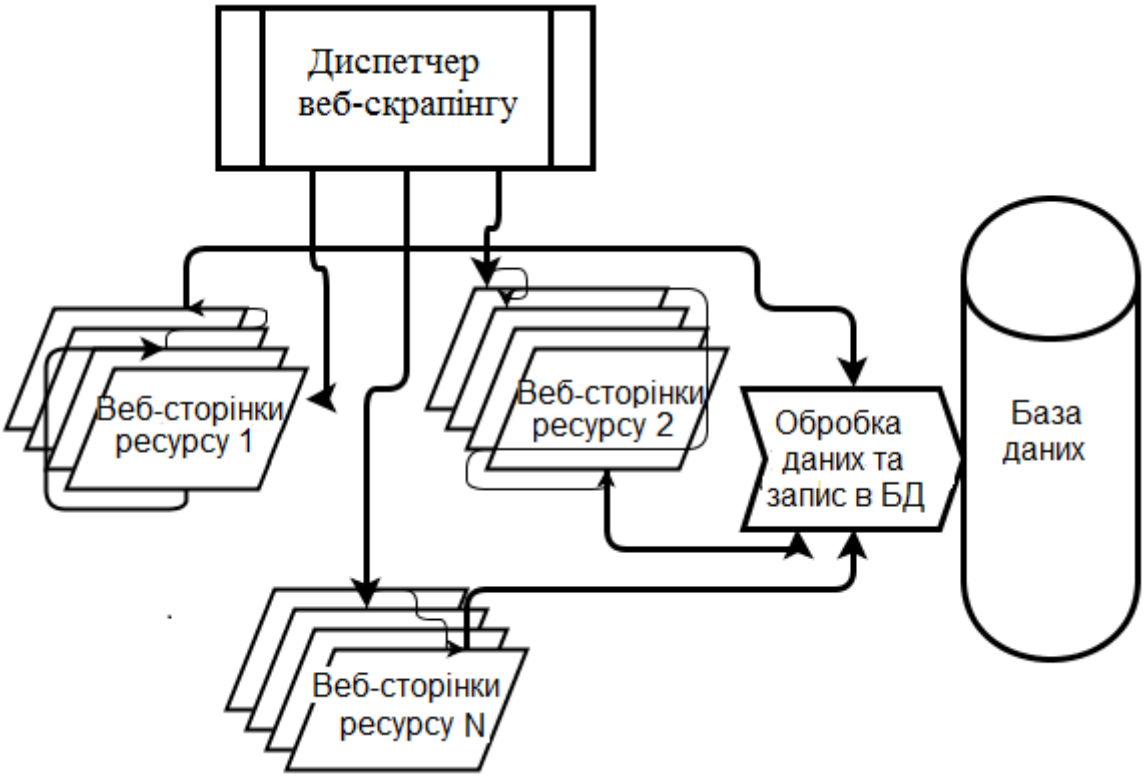


Рисунок 2.5 – Процес Веб-скрапінгу

Веб-скрапінг, також використовує програми типу павук і краулер для обходу і завантаження Веб-сторінок. На відміну від пошукових машин, сканується вузьке коло Веб-сторінок, заданих початковими умовами.

Після вилучення інформації в структурованому вигляді можлива подальша її обробка, яка може включати в себе фільтрацію результатів по деяким критеріям, підрахунок різних коефіцієнтів і показників, а також найбільш важливе і складне завдання – визначення авторів з однаковими прізвищами і, відповідно, підвищення точності результатів.

Виходячи зі змісту отриманої інформації, можна вирішувати проблему авторів з однаковими прізвищами декількома способами або їх комбінацією:

- семантичний аналіз теми або напрямки публікації;
- аналіз ключових слів публікації;
- аналіз джерела публікації.

В даному дослідженні спроектована система вилучення інформації про наукові публікації по параметру пошуку «Автор». Використовуючи цю властивість, програма виконує пошук по відомим їй наукометричних базах даних і завантажує результати та витягує інформацію певної структури.

На даний момент підтримуються доступ до наступних широко відомих міжнародних наукометричних баз даних:

- Scopus – бібліографічна і реферативна база даних та інструмент для відстеження цитованості статей, опублікованих в наукових виданнях. Позиціонується видавничою корпорацією Elsevier, як найбільша в світі універсальна реферативна база даних з можливостями відстеження наукової цитованості публікацій [81];
- Російський індекс наукового цитування (РИНЦ) – бібліографічна база даних виконує функцію не тільки інструменту для оцінки вчених або наукових організацій на основі цитування, але й авторитетного джерела бібліографічної інформації по науковій періодиці [82];
- BASE (Bielefeld Academic Search Engine) – багатопрофільна пошукова система для наукових інтернет-ресурсів, створена бібліотекою університету Біле-

фельд, Німеччина; є однією з найбільших пошукових систем публікацій в світі, особливо для відкритого академічного доступу до Веб-ресурсів [83];

– Index Copernicus – інтерактивна база даних з внесеної користувачем інформації про вченій профілі, наукових установ, публікацій і ін.; база даних має кілька інструментів оцінки продуктивності, які дозволяють відслідковувати вплив наукових робіт і публікацій, окремих вчених або науково-дослідних установ; також Index Copernicus пропонує традиційне реферування та індексування наукових публікацій. [84];

– Springer – міжнародна видавнича компанія, що спеціалізується на виданні академічних журналів та книг за природничо-науковими напрямами (теоретична наука, медицина, економіка, інженерна справа, архітектура, будівництво і транспорт); є другим за величиною видавництвом в світі після Elsevier в області «STM» (science, technology, medicine – англ. Наука, технології, медицина) [85].

Нижче розглянуті використовувані технології і засоби, використані при реалізації системи вилучення неформалізованій інформації з Веб-сторінок.

Використовуваний формат витягнутих даних – текстовий формат обміну даними (JSON). Структура складається з декількох полів, таких як «Автор», «Назва (публікації)», «Джерело», «Дата», наукометрична база та ін. Структура не жорстка, може відрізнятися набором полів для різних результатів, але, такі поля, як «Автор» і «наукометрична БД» є обов'язковими.

Структура даних має наступний вигляд:

```
{ "title" : "Informational Model of Natural Language Processing",
  "url" : "http://hdl.handle.net/10525/263",
  "author" : [
                "Palagin, Aleksandr",
                "Gladun, Viktor",
                "Petrenko, Nikolay",
                "Velychko, Vitalii",
                "Sevruk, Aleksey",
                "Mikhailyuk, Andrey"
            ],
  "spider" : "base-search",
```

```
  "source" : "Institute of Information Theories and Applications FOI ITHEA",
  "date" : "2008",
  "desc" : "The formal model of natural language processing in knowledge-based information systems is considered. The components realizing functions of offered formal model are described."}
```


Реалізація завантаження Веб-сторінок, навігація по посиланнях і вилучення даних з Веб ресурсів проводиться за допомогою Веб-скрапінг фреймворку Scrapy [86]. Вилучені дані зберігаються в NoSQL базі даних MongoDB [87], тому що вони не мають жорстких зв'язків, як в реляційних базах даних.

Фреймворк Scrapy надає зручний спосіб розширення числа підтримуваних наукометричних баз даних шляхом додавання нової програми-павука орієнтованого на роботу з Веб-ресурсом конкретної бази даних.

Використані технології і програмне забезпечення дозволяють створити програмний продукт по вилученню інформації з неоднорідних і неформалізованих джерел (таких як наукометричні бази) з перетворення її в структурований вигляд з можливою подальшою обробкою. Ці дані необхідні в першу чергу аспірантам і здобувачам при підготовці до захисту дисертацій. Крім того пропонована система може бути корисна при оцінці діяльності ВНЗ [51].

Розміщення публікацій в міжнародних наукометричних базах може мати позитивні наслідки для науки України. На прикладі бази Scopus на сайті Національної бібліотеки України ім. В. І. Вернадського показано, яку інформацію можна отримати: рейтинг вчених України, рейтинг організацій Національної академії наук України, рейтинг вищих навчальних закладів України і ін.

Представлений спосіб вилучення інформації з міжнародних наукометричних баз даних є свого роду універсальним інтерфейсом для програмного доступу до їх вмісту (хоч і обмеженому) [78]. Процес Веб-скрапінгу дозволяє вилучити неформалізовані даних з подальшим їх структуруванням. Для пошуку своїх публікацій, автору потрібно ввести своє прізвище та запустити програму. Далі результати в структурованому вигляді зберігаються в локальну (щодо наукометричних баз) базу даних і готові до подальшої обробки або перегляду.

2.4 Труднощі отримання даних з Веб сторінок і способи їх вирішення

Описана модель вилучення інформації опускає сам процес завантаження Веб сторінок з віддаленого сервера. На практиці цей процес може бути нетривіальним, що вимагає додаткової обробки. Зазвичай, процес отримання Веб сторінки складається з наступних кроків:

1. Клієнт посилає запит Веб серверу.

2. Веб сервер, як результат роботи, генерує Веб сторінку і відправляє її клієнту.

У більшості випадків, результуюча Веб сторінка – це HTML сторінка, яку клієнт обробляє і «вручну» витягує потрібну інформацію. Але також існують Веб сторінки, які крім HTML розмітки, містять певну частину програмного коду, який клієнт повинен інтерпретувати і виконати, щоб отримати кінцевий результат. Такий програмний код може містити в собі звернення до сервера за додатковою інформацією або динамічно створювати різні фрагменти HTML сторінки. У зв'язку з цим, обробка Веб сторінки в початковому вигляді від Веб сервера ускладнюється. Інформація може бути явно не представлена в результуючій Веб сторінці, а генеруватися на етапі виконання програмного коду, який міститься на цій сторінці.

Ще однією перешкодою до автоматичного вилучення даних з використанням Веб інтерфейсу є можливе блокування доступу клієнта до Веб сервера. При цьому Веб браузер, який запущений з того ж адресу клієнта, може мати доступ до запитуваної сторінці.

Таким чином, для досягнення максимально можливих результатів вилучення даних з Веб сторінок, потрібно вирішити наступні проблеми:

- обробка програмного коду, який присутній на Веб сторінці;

– надати можливість ідентифікації клієнта Веб сервером, щоб не бути заблокованим.

Обидві проблеми пов'язані з тим, що інформація з Веб сервера запитується без допомоги програми Веб браузера, для яких, вони в першу чергу призначені.

Але використання Веб браузера ускладнює автоматизацію процесу вилучення, вимагає залежності від зовнішньої програми, уповільнює роботу в цілому, так як браузер може мати зайвий функціонал, який не потрібен для отримання інформації.

Максимально можлива імітація роботи Веб браузера для завантаження Веб сторінки та її обробки (наприклад, виконання програмного коду) допоможе подолати зазначені труднощі. Проблема імітації роботи Веб браузера не нова, тому вже існують її рішення – використання так званих "безголових" браузерів (англ. Headless browser).

"Безголовий" браузер – це браузер без графічного інтерфейсу користувача. Вони забезпечують автоматизоване управління Веб-сторінками в середовищі, аналогічно до популярних Веб-браузерів, але виконуються за допомогою інтерфейсу командного рядка або за допомогою зв'язку через мережу. Вони особливо корисні для тестування Веб-сторінок, оскільки вони можуть показувати і розуміти HTML як і звичайний браузер, в тому числі розташування елементів, сторінки, колір, вибір шрифт, виконання програмного коду JavaScript і AJAX.

Найбільш поширеним "безголовим" браузером є PhantomJS – скриптовий браузер, який використовується для автоматизації взаємодії з Веб-сторінками. PhantomJS надає програмний інтерфейс для використання його іншими програмами і заснований на ядрі Webkit, який використовують такі бразузери як Safari і Google Chrome.

У разі використання такого браузера для завантаження Веб сторінки, можна спиратись на подальший спосіб обробки інформації, описаний в попередніх підрозділах. Таким чином, стає можливим вилучення інформації з різних Веб сторінок, в тому числі і тих, що динамічно конструюються на стороні клієнта.

2.5 Висновки до розділу 2

Майже кожна наукометричними база даних має різну структуру і різний спосіб зберігання інформації. Єдиного програмного інтерфейсу для автоматизованого доступу до них не існує. Але багато НМБД надають Веб інтерфейс у вигляді Веб сторінок для перегляду вмісту по заданому критерію (найчастіше ПІБ автора). Так як інформація на Веб сторінці в основному визначена мовою розмітки HTML, існують кілька способів обробки її:

– робота з Веб сторінкою як набором символів (наприклад, застосування регулярних виразів);

– використання мов запитів для слабоструктурованих даних (XPath – мова запитів до xml подібних документів);

– побудова і робота з об'єктною моделлю документа (DOM) – імітація роботи Веб браузера.

Використання перерахованих технік для отримання даних з Веб сторінок називається Веб скрапінгом. Комбінація різних способів дозволяє витягувати інформацію практично з будь-якої Веб сторінки, незалежно від її структури і вмісту.

Для вирішення завдання отримання даних з наукометричних баз даних пропонується модель програмного забезпечення, яка складається з так званих «програм-павуків», які обходять Веб сторінки певної бази даних і витягають метаданих публікацій по заданому критерію. Такий підхід використовують також пошукові машини (Google, Yandex), тільки в більшому масштабі – для сканування всіх сторінок доступних в мережі інтернет. Труднощі в основному виникають під час обробки динамічних Веб сторінок, які містять програмний код, який виконується на стороні клієнта. Для вирішення цієї проблеми використовується так званий «безголовий браузер», який завантажує і формує Веб сторінку як звичайний браузер, але без графічного інтерфейсу. При цьому надається програмний доступ до вмісту сторінки, що як раз і потрібно для автоматизованої обробки.

3 ТЕХНІЧНІ ПРОБЛЕМИ УПРАВЛІННЯ ІНФОРМАЦІЙНИМИ СЕРЕДОВИЩАМИ

3.1 Векторна парадигма методології управління проектами

Проектно-векторний простір (ПВП) – це простір, утворений в системі координат, що визначають можливі стани організаційних, методологічних, технологічних і продуктових компонентів проектів, що реалізуються в освітніх середовищах [88, 89]. Атрибути проектно-векторного простору – координати (простір-утворюючі категорії) і його наповнення.

Формально проектно-векторний простір можна представити як кортеж

$$\Omega = N_1 \times N_2 \times ... \times N_i \times ... \times N_p,$$

де N_i – вимірювання проектно-векторного простору;

Ω – проектно-векторний простір.

Наповнення ПВП формує «речовину, енергію та інформацію». Аналогом енергії фізичного простору в проектно-векторному просторі виступають гроші, а інформація є його основним атрибутом. Аналогом речовини виступають об'єкти і суб'єкти проектів, що формують проектно-векторне середовище (проекти, продукти, інструменти та суб'єкти) проектів.

Об'єктами проектно-векторного простору (об'єктами проектів) є відокремлені певним поняттям сутності, що відноситься до ресурсів, продуктів або інструментів і впливають на процеси в проектах.

Суб'єкти проектів формують наповнення ПВП, що змінюється в процесі реалізації проектів. Суб'єкти ПВП є джерелом і носієм відношення до того, що відбувається в проекті. Це менеджери, виконавці, вище керівництво, зацікавлені сторони. Суб'єкти проектно-векторного простору (суб'єкти проектів) - представники юридичних осіб або фізичні особи, зацікавлені в реалізації (або в нереалізації) проекту або в отриманні продукту проекту і виражають своє ставлення до об'єктів ПВП через їх суб'єктивну оцінку та оцінку їх розвитку.

До суб'єктів проектів відносяться люди, які що-небудь одержують від проекту, або що-небудь віддають у проект. Це керівники або інші учасники проектів, функціональні та проектні менеджери, виконавці. Для оцінки розвитку ПВП важливо не тільки розуміння ролі суб'єктів проектів, але також розуміння їх потреб, вимог до компетенцій, і, відповідно, цілей, що стоять перед ними.

Цілі проектів – створюваний потребою суб'єктів проектів прийнятний орієнтир розвитку освітніх середовищ. Зміни в об'єктах ПВП націлені на отримання деякого об'єкта, ототожнюються з метою проекту – продуктом проекту.

Продукт проекту – матеріально-технічний або інформаційний об'єкт ПВП, створюваний у процесі реалізації проекту, що задовольняє потреби зацікавлених сторін проекту.

Інструменти проектів – об'єкти проектів, використовувані в процесі створення продукту проекту. До інструментів проектів відносяться методи, засоби, обладнання, матеріали, ресурси, інформація, використовувані в процесі створення продукту проекту.

Атом об'єкта проектно-векторного простору – його абстрактний або реальний елемент об'єкта ПВП, що має ті ж координати в проектно-векторному просторі, що й об'єкт, і не вимагає поділу на складові частини (на час реалізації проекту).

Атом суб'єкта проектно-векторного простору – понятійний елемент суб'єкта ПВП, що відображає тільки одне ставлення до стану або розвитку проекту.

Макроутворення проектно-векторного простору – постійні на деякому відрізку часу сукупності об'єктів і суб'єктів ПВП, які виділяються для підвищення ефективності управління проектами.

Вимірювання проектно-векторного простору характеризуються тим, що чим більше значення координати, тим більш розвиненою є об'єкт чи суб'єкт, проекція якого на цю вісь має дане значення координати. Тому, можна сказати, що кожна координата містить всі координати, значення яких менше даної. Це означає, що сума інвестицій в 1 млн грн. містить інвестиції з сумою 100000 грн, а 50% виконання робіт містить 30% виконання робіт.

Проектно-векторний простір зручно розглядати як дискретний. Це означає, що в дискретні відрізки часу здійснюються стрибкоподібні переміщення об'єктів проектів. Кроком дискретизації може бути день, тиждень, місяць, рік залежно від масштабності проекту або проектів, що утворюють цей простір.

Проектно-векторний простір в розрізі реалізованих проектів можна уподібнити «розширюваному Всесвіту» [88]. Спочатку в проекті нічого немає (точніше, проекту немає) і простір згорнуто в точку. У міру того, як у розрізі вимірювань формуються нові об'єкти і суб'єкти проекту простір починає розширяться (рис. 3.1).

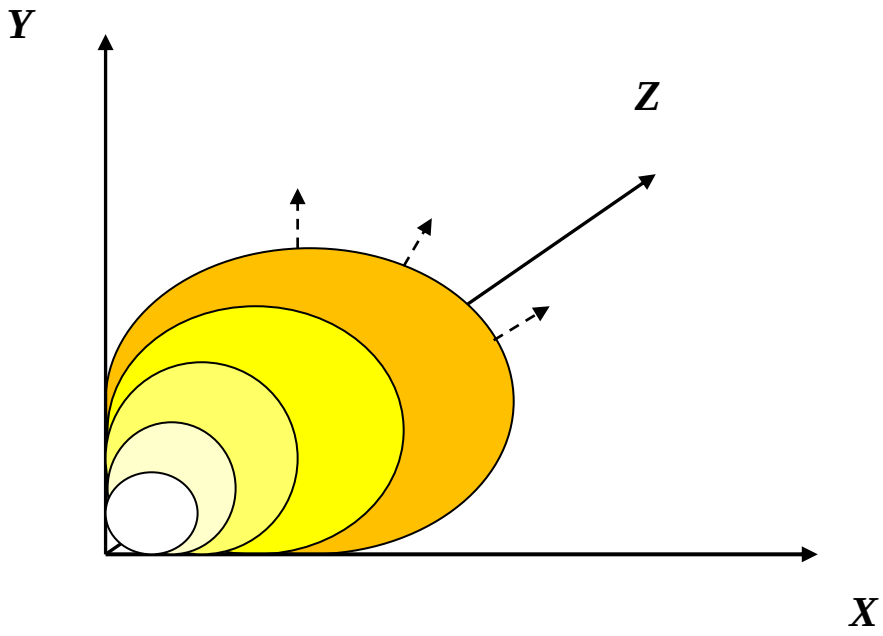


Рисунок 3.1 - Проект як «розширюваний Всесвіт» в проектно-векторному просторі

Під проектним вектором будемо розуміти сутність, що реалізується в проектній діяльності та ідентифікується сукупністю координат, напрямок зміни об'єктів і суб'єктів проекту. Проекція вектора на вісь (осі) координат відображає індивідуальні особливості проекту в розрізі компонентів проектів освітніх середовищ, що сполучені з цими осями координат.

Підхід до побудови системи управління проектами освітніх середовищ, який базується на виділенні і оптимізації векторів у ПВП будемо називати про-

ектно-векторним підходом до управління освітніми середовищами. Модель проектно-векторної системи управління освітніми середовищами можна представити сукупністю векторів, кожен з яких визначає зміни об'єктів і суб'єктів проекту.

Систему управління проектами освітніх середовищ, що реалізовує проектно-векторний підхід будемо називати системою проектно-векторного управління освітніми середовищами (СПВУОС). Для побудови такої системи необхідно спочатку розробити методологію проектно-векторного управління освітніми середовищами. Методологія проектно-векторного управління освітніми середовищами - система понять, методів, методик, структур і засобів їх реалізації в організації та управлінні проектами, в основі якої лежить проектно-векторний підхід до управління освітніми середовищами.

У векторній парадигмі проглядається дві основні переваги перед іншими концепціями на створення систем управління.

По-перше, це декомпозиція досить складною організаційно-технічної системи організацій, що належать до освітніх середовищ на прості, орієнтовані на розвиток окремих об'єктів і суб'єктів проектів компоненти, описувані проектно-інформаційними, проектно-процедурними і проектно-технологічними векторами.

По-друге, до багатьох видів діяльності ООС (здійснюваних не тільки в традиційних проектах) можна застосувати проектно-векторний підхід. А це дозволяє використовувати досить потужний інструмент управління проектами для удосконалення процесів управління організаціями в освітніх середовищах.

В основі проектно-векторного підходу до управління освітніми середовищами лежить подальший розвиток ідей, методів і моделей, які розроблені в рамках наукових основ матричних інформаційних технологій [89] і матричних технологій управління. Насправді інструменти, що використуються для реалізації різних проектів не двоорієнтіровани, як в матричних технологіях, а орієнтовані відповідно до структури продуктів проектів і змістом тих методів і засобів, які і забезпечують реалізацію інформаційних проектів.

Проектно-векторний підхід на відміну від матричних (двокомпонентних) технологій управління проектами являє собою n-компонентну структуру (кожен компонент являє собою один вимір проектно-векторного простору), яка базується на різних подмножествах методів, засобів управління і реалізації проектів, але в сукупності представляють собою єдиний, хоча і разноорієнтірованних процес розвитку як внутрішнього середовища, так і продуктів проектів (рис. 3.2).

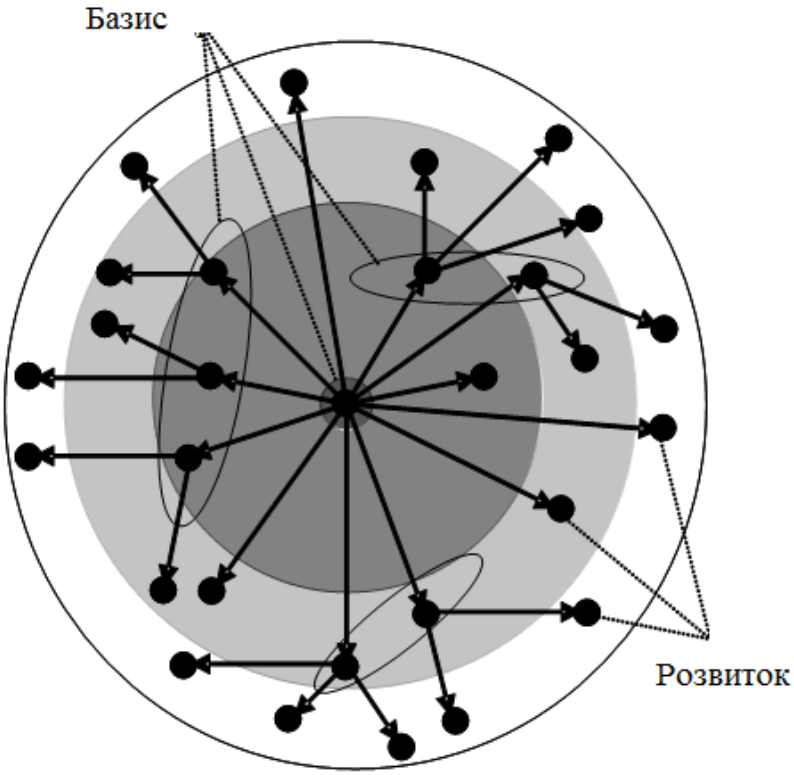


Рисунок 3.2 – Двумірне відображення проектно-векторного простору

Множина векторів проектно-векторного простору відповідає множині об'єктів цього простору і задається функцією:

$$A_k^{(j)} = \varphi(\Omega, \Gamma)$$

де $A_k^{(j)}$ – *j*-й () вектор для *k*-го проекту;

$\varphi(...)$ – функція, що задається алгоритмічно;

Ω – проектно-векторний простір;

Г – об'єкти і суб'єкти традиційних, операційних та управлінських проектів (об'єкти і суб'єкти ПВП).

Кожен вектор задається координатами, обумовленими часом розвитку деякого об'єкта / суб'єкта проекту в проектному просторі:

$$A_k^{(j)}(t) = \left[\mathrm{x}_{\mathrm{k1}}^{(\mathrm{j})}(\mathrm{t}), \mathrm{x}_{\mathrm{k2}}^{(\mathrm{j})}(\mathrm{t}), \ldots, \mathrm{x}_{\mathrm{ki}}^{(\mathrm{j})}(\mathrm{t}), \ldots, \mathrm{x}_{\mathrm{kp}}^{(\mathrm{j})}(\mathrm{t})\right],$$

де $\mathrm{x}_{\mathrm{ki}}^{(\mathrm{j})}(\mathrm{t})$ – значення координати об'єкта / суб'єкта Q_j проекту $П_k$ по осі N_i в проектно-векторному просторі в момент часу t.

Математично система проектно-векторного управління освітніми середовищами (СПВУОС) повинна відображати сформовані в проектно-векторному просторі вектори (напрям зміни об'єктів), оцінювати і коригувати їх, виходячи з потреб суб'єктів проектів та їх цілей. Оцінка ефективності СПВУОС повинна здійснюватися через оцінку відстані між векторами, що відображають потрібний і фактичний розвиток об'єктів і суб'єктів проектів.

3.2 Інструменти забезпечення управління інформаційними проектами

В даний час графічні моделі є основним інструментом для побудови імовірнісних тематичних моделей (probabilistic topic model). Моделі з прихованими змінними виявилися особливо ефективними для виявлення прихованих структур в текстових колекціях. Важливим є підклас орієнтованих імовірнісних тематичних моделей (directed probabilistic topic models, DPTM), які здійснюють м'яку кластеризацию і застосовуються для виявлення тематики текстів у великих колекціях документів. У термінах кластерного аналізу тема (topic) – це результат бі-кластеризації, тобто одночасної кластеризації і слів, і документів по їх семантичної близькості. При м'якій кластеризації (soft clustering) кожне слово і кожен документ належить до кількох тем одночасно з певними можливостями. Таким чином, стислий семантичний опис слова або документа є імовірнісний розподіл на безлічі тем. Процес знаходження цих розподілів і називається тематичним моделюванням.

Графічні моделі можуть бути розділені на дві основні категорії: орієнтовані і неорієнтовані графічні моделі. Ці типи можна далі розбити на параметричні і непараметричні.

Графічні моделі з прихованим тематичним шаром останнім часом успішно застосовувалися в задачах пошуку прихованих закономірностей у даних. Автоматичне виділення тематики текстів застосовувалося для розбиття текстів по групах на основі семантичної близькості змісту. Ці моделі дозволяють класифікувати документи, але вони обмежені припущенням, що кожен документ стосується тільки одного кластеру. Моделі м'якої кластеризації дозволяють відносити документ одночасно до кількох кластерів, при цьому кожен кластер асоціюється з певною темою, і кожен документ характеризується оцінками близькості до кожної з тем. В орієнтованих імовірнісних тематичних моделях (directed probabilistic topic model, DPTM) оцінка близькості документа до теми має імовірнісний сенс і може інтерпретуватися як частка вмісту документа, що відноситься до даної теми. DPTM – це відносно молода галузь досліджень в теорії самонавчання (навчання без учителя, unsupervised learning), що представляє в даний час значний як теоретичний, так і практичний інтерес. Одним з перших був запропонований імовірнісний латентний семантичний аналіз (probabilistic latent semantic analysis, PLSA, рис. 3.3), заснований на принципі максимуму правдоподібності, як альтернатива класичним методам кластеризації, заснованим на обчисленні функцій відстані.

Імовірнісний латентний семантичний аналіз (probabilistic latent semantic analysis, PLSA) – заснований на введенні шару прихованих змінних для опису тематик в колекції текстових документів. Модель PLSA є важливою віхою в розвитку імовірнісного моделювання текстів, і вона, поза сумнівом, корисна для завдань інформаційного пошуку. Однак вона має досить суттєві обмеження. У PLSA кожний документ відображається числовим вектором, кожна компонента якого містить частку відповідної теми в документі. Однак імовірнісна модель не описує ні закон розподілу цих часток, ні ймовірності самих документів. В результаті число параметрів моделі лінійно зростає з ростом розміру текстової

колекції, що може призводити до перенавчання. Крім того, не зрозуміло, як оцінювати ймовірності нових документів, що не входили до складу навчальної вибірки.

Тематичні моделі активно розвивалися останнім десять років. Запропоновано багато моделей для вирішення задач моделювання текстових колекцій в таких додатках, як класифікація документів, пошук схожих документів, пошук експертів, виявлення спільнот і аналіз тимчасових трендів.

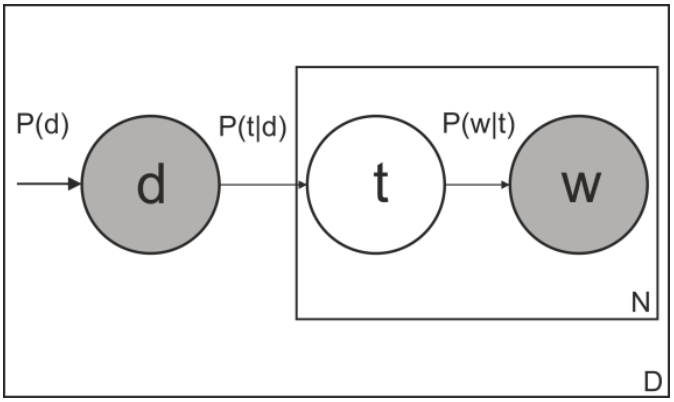


Рисунок 3.3 – Графічна імовірнісна модель PLSA:

d – документ; w – слово (d, w – спостережувані змінні);

t – тема (прихована змінна);

p(d) – апріорний розподіл на множині документів;

p(w|t), p(t|d) – шукані умовні розподілу;

D – колекція документів;

N – довжина документа в словах.

Слідом за PLSA був запропонований метод латентного розміщення Дирихле (latent Dirichlet allocation, LDA) і його численні узагальнення. Застосування тематичних моделей дозволяє отримати відповідь на цілий ряд нетривіальних питань:

Як виявляти сенс або тематику документів за їх змістом?

Як здійснювати класифікацію документів на основі цих прихованих тематичних закономірностей?

Як виявляти наукові інтереси авторів і знаходити експертів в спеціальних областях знання?

Як виявляти приховані асоціативні зв'язки між окремими дослідниками або групами людей?

Як підбирати колаборації під проекти?

Як виявляти тенденцій у розвитку наукових напрямків?

Як виявляти ролі людей в соціальних мережах?

Як здійснювати індексацію і автоматичне анотування документів?

Під параметрическими моделями будемо розуміти тематичні моделі, в яких теми спочатку фіксовані і не змінюються в процесі побудови моделі. У непараметрических моделях число тим спочатку не фіксоване, а самі теми налаштовуються в процесі пошуку найкращого можливого модельного опису даних.

Нижче описуються концепції і термінологія, що лежать в основі теорії тематичного моделювання.

Документ зазвичай складається з множини слів, термінів (словосполучень), пеціальних символів, таблиць, ілюстрацій, і ін. У дослідженнях з тематичного моделювання типовими видами документів є наукові статті та новинні повідомлення.

Велику колекцію текстових документів прийнято називати корпусом текстів (text corpora). У дослідженнях по тематичному моделювання часто використовують загальнодоступні корпуси «NIPS proceedings» і «Cite seer». Обидва містять велику кількість наукових статей і використовуються для тестування різних методів виявлення знань. Відомі корпуси «TREC AP» і «Reuter's» використовуються для тестування методів аналізу новин.

У літературі по тематичним моделям поняття *теми* (topic) визначається порізному, в залежності від наукової школи: «приховані патерни», «компактні опису сенсу документів», «ймовірні (нечіткі) кластери семантично пов'язаних термінів», «сполучна ланка між термінами і іншими об'єктами (документами,

авторами, організаціями, конференціями, і ін.), яке дозволяє знаходити приховані асоціативні зв'язки між ними ».

Формально тема визначається як дискретне (поліноміальний) імовірнісний розподіл в просторі слів заданого словника. Документ може складатися з величезного числа слів, однак ці слова можуть породжуватися невеликим числом тим, як сумішшю поліноміальний розподіл.

Припущення про те, що для цілей аналізу текстів (в нашому випадку – для виявлення тематики) важлива тільки частота слів, але не їх порядок, називається *моделлю мішка слів* (bag of words). Коли не важливий порядок пропозицій в документі або порядок документів в корпусі, відповідним чином вводяться моделі мішка пропозицій і мішка документів.

Основна ідея тематичного моделювання полягає в описі документа сумішшю тим, тобто у визначенні документа як вибірки слів, яку породжує сумішшю імовірнісних розподілів.

Тематичним моделюванням називається рішення оберненої задачі. Кожен документ в корпусі текстів розглядається як спостережувана випадкова незалежна вибірка слів (мішок слів), породжена деяким, як правило невеликим, латентним підмножиною тем. За цими даними потрібно відновити імовірнісні розподілу всіх тем в корпусі і визначити, яким саме підмножиною тим породжений кожен документ.

3.3 Інструментрій латентно семантичного аналізу для ідентифікації схожих публікацій

Проблема ідентифікації текстів природної мови обчислювальними машинами давно представляє науковий інтерес [80]. В даний час широко використовуються різні підходи розпізнавання мови, класифікації текстової інформації, визначення ідентичності текстів [90 – 92]. Автоматизація добування інформації з наукометричних баз даних пов'язана з необхідністю уточнення результатів запитів до баз даних в частині виключення «двійників» авторів, у яких збігаються

прізвища та ініціали. Подібні проблеми виникають і при розробці систем автоматизованого навчання с відкритими тестами [93 – 96].

Одним із прикладів вилучення основного змісту з текстів є пошук схожих документів або документів з певної тематики. Стандартний пошук використовує порівняння документів на наявність шуканого рядка або слів. Однак не завжди можна сформулювати точний запит. Часто потрібен пошук, який заснований на аналізі смислового навантаження документів. Одним з підходів, який вже активно використовують пошукові гіганти, є латентно семантичний аналіз. Цей підхід дозволяє виявити закономірності у відносинах між поняттями і термінами в неструктурованою колекції текстів.

Існує кілька способів смислового аналізу текстів, які можна розділити на наступні групи [80]:

- лінгвістичний аналіз;
- статистичний аналіз.

Перша група орієнтована на визначенні смислу по семантичній структурі тексту і включає лексичний, морфологічний, синтаксичний і аналіз. В даний час відсутні сформовані підходи до реалізації завдання семантичного аналізу текстової інформації, що багато в чому обумовлено винятковою складністю проблеми і недостатньо повної опрацюванням наукового напрямку створення систем штучного інтелекту.

Друга група – це, як правило, частотний аналіз в тих чи інших його варіаціях. Сутність аналізу полягає в підрахунку кількості повторень слів в тексті і використанні результатів підрахунку для конкретних цілей. Всілякі варіанти різних реалізацій підрахунку слів і подальша обробка результатів підрахунку утворюють широкий спектр пропонованих в даному класі методів і алгоритмів.

Одним з найбільш ефективних статистичних підходів є латентно семантичний аналіз (або латентно семантичне індексування) [72]. Автори представили модель дворежимного факторного аналізу, яка заснована на сингулярному розкладанні (SVD). Сингулярне розкладання представляє терміни і документи у

вигляді векторів в просторі обраній розмірності, а скалярний добуток між точками простору – їх схожість.

Латентно семантичний аналіз починається з побудови матриці документів і термінів – індексованих слів [92]. Індексовані слова це слова, які зустрічаються в двох або більше документах і мають смислове навантаження (не є сполучники і прийменники і ін.). Далі застосовується сингулярне розкладання цієї матриці на добуток трьох матриць:

$$A = U \cdot S \cdot V^t,$$

де матриці U та V – ортогональні, а S – діагональна матриця, на діагоналі якої значення називаються сингулярними значеннями матриці A.

Таке розкладання володіє чудовою особливістю: якщо в матриці S залишити тільки k найбільших сингулярних значень, а в матрицях U і V – тільки відповідні цим значенням стовпці, то добуток матриць S, U і V буде найкращим наближенням вихідної матриці A до матриці $\hat{A}$ рангу k :

$$\hat{A} \approx A = U \cdot S \cdot V^t$$

Основна ідея латентно-семантичного аналізу полягає в тому, що якщо в якості матриці A використовувалася матриця індексованих слів для документів, то матриця $\hat{A}$, що містить тільки k перших лінійно незалежних компонент A, відображає основну структуру різних залежностей, присутніх у вихідній матриці. Структура залежностей визначається ваговими функціями індексованих слів. Таким чином, кожне индексоване слово-терм і документ представляються за допомогою векторів в загальному просторі розмірності k. Близькість між будь-якою комбінацією індексованих слів і / або документів легко обчислюється за допомогою скалярного добутку векторів [14]. Як правило, вибір залежить від поставленого завдання і підбирається емпірично. Якщо вибране значення занадто велике, то метод втрачає свою потужність і наближається за характеристиками до стандартних векторних методів. Занадто мале значення k не дозволяє вловлювати відмінності між схожими термами або документами.

Латентно-семантичний аналіз добре долає проблему синонімії, але частково з проблемою полісемії, тому, що кожне слово визначається однією точкою в

просторі. Також цей аналіз дозволяє виконувати автоматичну категоризацію документів, засновану на схожості їх концептуального змісту. Також перевагою латентно семантичного аналізу є незалежність від мови, так як це математичний підхід. Недоліком методу є зниження швидкості обчислення при збільшенні обсягу вхідних даних (наприклад, при SVD-перетворенні).

Вилучення інформації з наукометричних баз даних потребує постобработке з метою визначення схожих за змістом публікацій, а також визначення дублікатів. Мета даного дослідження розробити спосіб семантичного аналізу витягнутої інформації.

Застосування латентно семантичного аналізу для проекту з вилучення інформації з наукометричних баз даних дозволить розділити отримані публікації на категорії з метою визначення однофамільців. Наприклад, автор Іванов І. І. займається дослідженнями в області комп'ютерних наук, але результати пошуку його публікацій в наукометричних базах містять багато невідповідних записів, так як є ще один автор Іванов І.І., який опублікував статті з медичної тематики. Латентно семантичний аналіз дозволить розділити публікації, які відносяться до різних концептів. Різні наукометричних баз можуть містити дублікати публікацій в дещо зміненій формі. Визначення цих дублікатів також можливо за допомогою латентно семантичного аналізу.

Розглянемо послідовність дій латентно семантичного аналізу, зображену на рис. 3.4, до деякого набору документів. В якості прикладу візьмемо невеликий список назв публікацій, витягнутих з наукометричних баз даних для автора Колесникова Є. В. (табл. 3.1). Вибір автора добре підходить для прикладу, так як існує кілька авторів з однаковим прізвищем та ініціалами.

Переглянувши список документів, можна помітити, що частина статей відноситься до медичної тематики, а частина – до управління проектами. Щоб провести цей поділ застосовується латентно семантичний аналіз.

Спочатку є список тем, який потрібно проаналізувати і обробити з метою виділення індексованих слів.

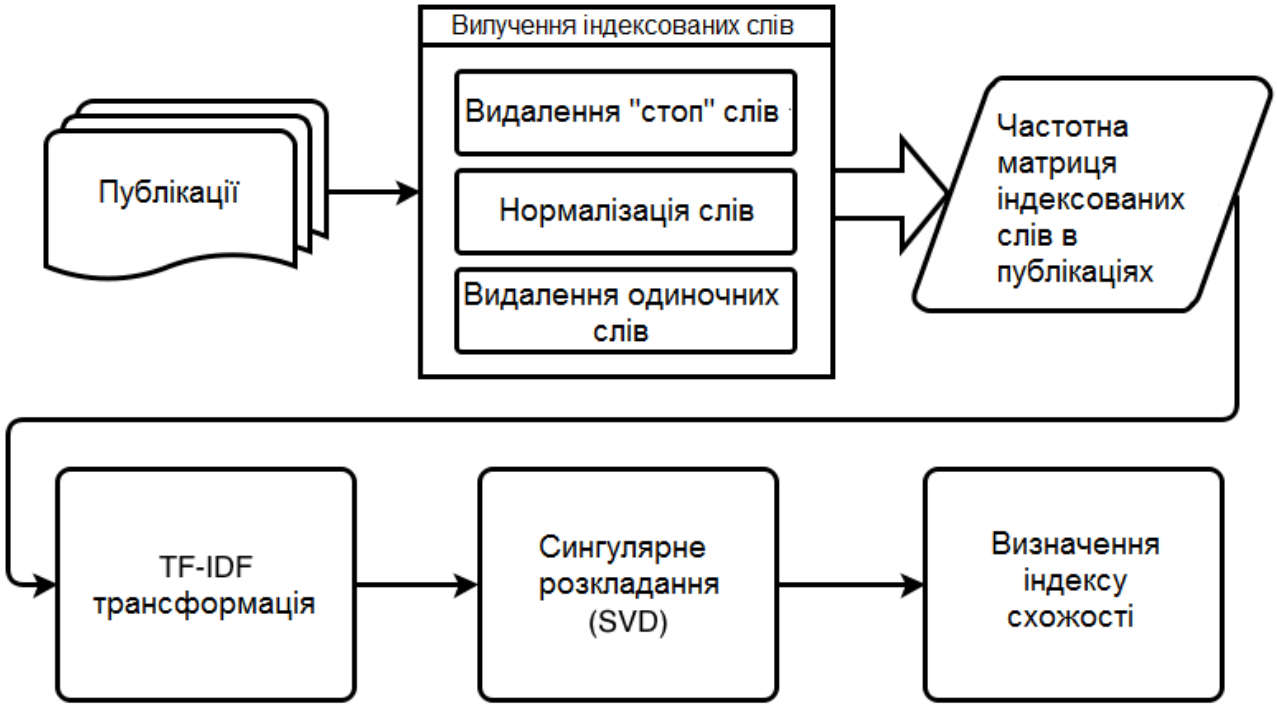


Рисунок 3.4 – Послідовність дій латентно семантичного аналізу

Таблиця – 3.1 Список документів для прикладу роботи латентно-семантичного аналізу

Д1	Когнітивні моделі слабо структурованих проектів створення програмних продуктів
Д2	Лікарсько індуковані ураження печінки: особливості виявлення, постановки діагнозу і ведення пацієнтів
Д3	Трансформація когнітивних карт в моделі марківських процесів для проектів створення програмного забезпечення
Д4	Особливості ураження печінки при ВІЛ інфекції
Д5	Матрична діаграма і сильна зв'язність індикаторів цінності в проектах
Д6	Патогенетичні механізми прогресування поєднаних вірусних і алкогольних уражень печінки
Д7	Розробка марковских моделей змін стану пацієнтів в проектах надання медичних послуг
Д8	Вирішені та невирішені питання терапії неалкогольний жирової хвороби печінки в рамках метаболічного синдрому
Д9	Аналіз структурної моделі компетенцій з управління проектами національного стандарту України

Аналіз виконується у такій послідовності:

– видалення, так званих, "стоп" слів, тобто, таких, які не мають смислового навантаження (прийменники, сполучники і т.д.);

– приведення слів до нормального вигляду або стемінг – процес знаходження основи слова (використовується алгоритм Портера [97], який дозволяє швидко визначити основу слова);

– видалення слів, що зустрічаються тільки один раз. Цей пункт не обов'язковий, але дозволяє економити ресурси при розрахунках.

На основі отриманих індексованих слів будується частотна матриця використання цих слів (табл. 3.2).

Таблиця 3.2 – Частотна матриця використання індексованих слів в документах за атрибутом: автор «Колесникова Е.В.»

Індексовані слова	Документи								
	Д1	Д2	Д3	Д4	Д5	Д6	Д7	Д8	Д9
когнитивн	1		1						
марковск			1				1		
модел	1		1				1		
особен		1		1					
пациент		1					1		
печен				1		1		1	
поражен		1		1		1			
программн	1		1						
проект	1		1		1		1		1
создан	1		1						

Для підвищення якості аналізу, виконується наступний етап – трансформація матриці за допомогою моделі TF-IDF (від англ. TF – termfrequency, IDF – inverse document frequency) – статистична міра, яка використовується для оцінки важливості слова в контексті документа, що є частиною колекції документів. Вага певного слова пропорційна числу вживання цього слова в документі, і обернено пропорційна частоті вживання слова в інших документах колекції.

Наступний крок, є основою латентно семантичного аналізу – це сингулярне розкладання отриманої матриці і побудова індексу схожості, який обчислюється за відстанню між індексованими словами і документами в *k*-вимірному просторі. На рис. 3.5 показано графічне представлення індексованих слів і заголовків в двовимірному просторі ($k = 2$), а таблиця 3.3 містить їх координати, отримані з сингулярного розкладання.

Слід підкреслити, що стратегія виконання зазначених вище завдань повинна будуватися на основі законів проектного управління [98]. При цьому необхідно враховувати загальні підходи до організації термінологічних систем наукового знання [99], моделі комунікаційних процесів [10], а також особливості організації комп'ютерних мереж [101].

Таблиця 3.3 – Координати слів і документів у двувимірному просторі

Индексоване слово	X	Y	Документ	X	Y
когнитивн	-0.33	0.05	Д1	-0.56	0.06
марковск	-0.28	-0.02	Д2	-0.04	-0.55
модел	-0.52	0.01	Д3	-0.64	0.05
особен	-0.01	-0.47	Д4	-0.01	-0.65
пациент	-0.12	-0.26	Д5	-0.15	0.00
печен	0.00	-0.52	Д6	-0.01	-0.47
поражен	-0.02	-0.66	Д7	-0.40	-0.10
программн	-0.33	0.05	Д8	0.00	-0.21
проект	-0.56	0.01	Д9	-0.29	0.01
создан	-0.33	0.05			

На рис. 3.3 показано, що частина документів належить до однієї тематики, а решта - до іншої. Проаналізувавши індексовані слова, що знаходяться поруч з документами, можна зробити висновок, що тематика 1 відноситься до управління проектами (проект, модел, програмн), а тематика 2 – до медицини (пацієнт, печен, поражен).

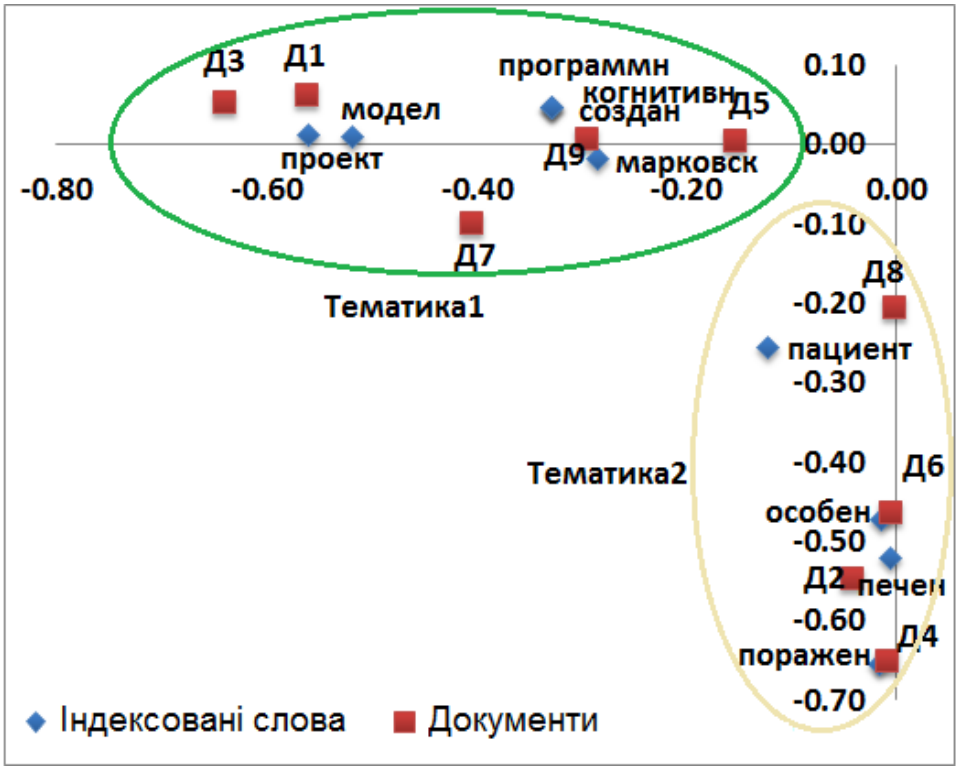


Рисунок 3.5 – Графічне представлення розподілу індексованих слів і документів в двовимірному просторі

Латентно семантичний аналіз надає досить непогані результати порівняння різних документів за змістом і дає можливість автоматичної категоризації їх. Будучи заснованим на математичних і статистичних розрахунках, цей підхід є незалежним від мови документів.

Застосування латентно семантичного аналізу в проекті добування інформації з наукометричних баз даних дозволяє вирішувати проблему авторства наукових публікацій і виявити дублікати.

3.3 Достовірність ідентифікації авторства наукових публікацій на основі латентно семантичного аналізу

Одним з етапів вилучення та збору інформації є її обробка. На відміну від технічного процесу вилучення, обробка може являти собою інтелектуальну і навіть творчу роботу. Основним завданням на цьому етапі є визначення досто-

вірності результатів. Прикладом може служити наступне завдання: задані прізвище, ім'я та по батькові (ПІБ) учасника та список публікацій, витягнутих з цих атрибутів; як визначити статті, тільки цього автора (оскільки атрибут запиту ПІБ для різних авторів може збігатися).

Латентно семантичний аналіз використовується в обробці природної мови, когнітивної науці і комп'ютерної лінгвістики для вирішення подібних завдань. Найбільш відомими практичними реалізаціями ЛСА на сьогоднішній день є:

– SenseClusters. Основна функція – кластеризація схожих контекстів. Застосовується при вирішенні неоднозначності слів (зокрема, імен), класифікації документів різного роду (електронних листів, новинних статей), класифікація лексики (знаходження синонімів, антонімів і інших класів відносин) [73].

– S-Space. Основна функція – універсальний засіб для побудови і обробки векторної моделі. Містить реалізації великої кількості алгоритмів (різні векторні моделі, деякі методи їх подальшої обробки). Орієнтоване на швидкість роботи, інтуїтивно зрозуміле уявлення даних [74].

– Gensim. Найбільш надійне і ефективне програмне забезпечення, яке реалізує семантичне моделювання для звичайного тексту. Призначено спеціально для обробки великих колекцій документів, з використанням ефективних алгоритмів [75].

Проект по вилученню інформації з наукометричних баз даних (НМБД) має на увазі отримання інформації про публікації, які належать конкретному авторові, з найбільш відомих НМБД [76]. Виконання пошуку по заданому аргументу – ПІБ – дозволяє отримати список публікацій, автором яких, по ідеї, є одна людина. Але це не завжди вірно, так як ПІБ автора не може бути унікальним ідентифікатором запису. У світі можуть існувати кілька авторів з однаковими ПІБ. Додамо до цього той факт, що найчастіше публікації містять тільки ініціали з прізвищем, тому ймовірність знаходження публікацій кількох авторів з ідентичними ПІБ, ще вище. Тому для вибірки публікацій належать одному автору, потрібно використовувати додаткову інформацію з доступних полів структури даних назва публікації. Назва може відображати напрямок діяльності ав-

тора, а також це обов'язкове поле, яке не може бути порожнім, в той час як інші поля часто не доступні в тих чи інших наукометричних базах [77]. На рис. 3.6 показаний алгоритм використання ЛСА для класифікації публікацій.

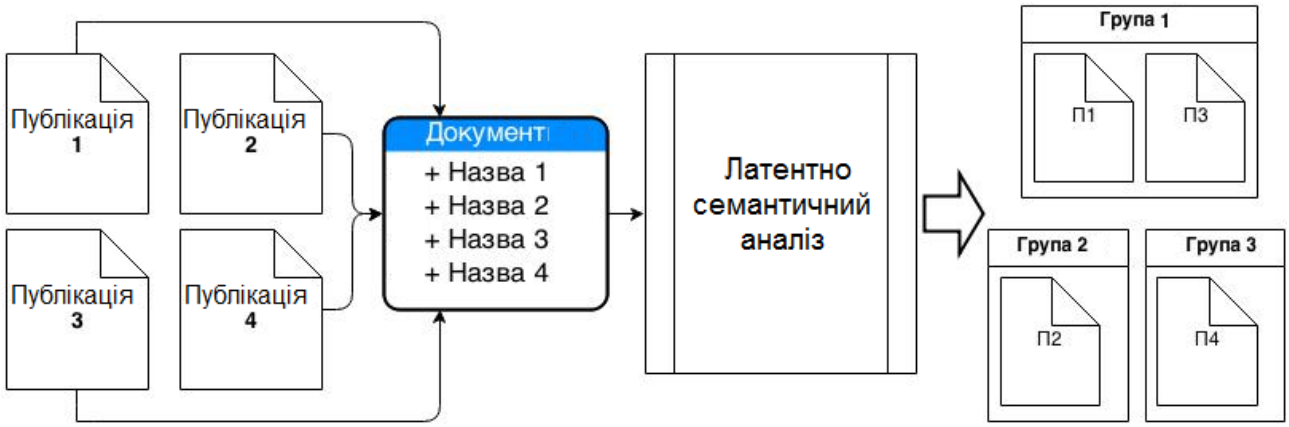


Рисунок 3.6 – Застосування латентно семантичного аналізу для визначення авторства наукових публікацій

Розглянемо застосування латентно семантичного аналізу при аналізі назв публікацій. У таблиці 3.4 представлений приклад результатів пошуку публікацій для автора "Колесникова Е. В".

Як видно, частина статей пов'язана з медичною тематикою, але останні дві публікації (позначені зірочками) відносяться до зовсім іншої предметної області. Якщо відомо, напрямок діяльності автора, то можна визначити з деякою погрішністю, які з публікацій належать даному автору. Для того, щоб цей процес автоматизувати, можна виділити ключові слова зі сфери предметної області автора і за допомогою програми відібрати підходящі варіанти. Але тут виникає проблема: нам потрібен набір з множини слів, які можуть зустрічатися в назвах статей. Цей набір може бути занадто об'ємним, що позначиться на продуктивності. Крім цього слід брати до уваги, що деякі слова можуть вживатися в різних контекстах з різним змістом (проблема полісемії).

Таблиця 3.4 – Фрагмент публикацій за запитом “Колесникова Е.В.”

№	НМБД	Название публикации
1	Base-search	Лекарственно-индуцированные поражения печени: особенности выявления, постановки диагноза и ведения пациентов
2	Base-search	Современное состояние проблемы самоубийств в судебной медицине
3	Base-search	К вопросу о патоморфологических исследованиях нейроэндокринной системы при завершенных суицидах
4	Base-search	Теоретические исследования рабочего цикла гидравлического устройства ударного типа для ликвидации прихватов бурового снаряда в разведочных скважинах
5	Base-search	Гипоадипонектимия- ключевой фактор риска неалкогольной жировой болезни печени (обзор литературы)
6	Base-search	Особенности диагностики при подозрении на диффузную форму рака молочной железы
7*	Base-search	Трансформация когнитивных карт в модели марковских процессов для проектов создания программного обеспечения
8*	Base-search	Развитие теории проектного управления: обоснование закона К. В. Кошкина о завершении проектов

Щоб вирішити ці проблеми використовується латентно семантичний аналіз, який дозволяє виділити зв'язок між ключовими словами і набором документів (назв публікацій, в нашому випадку). Припустимо, задано ключове слово "інформація". Застосування латентно семантичного аналізу дозволяє встановити приховані зв'язки, наприклад, слова "програма" або "комп'ютер" близькі до предметної області цього слова. Тому ЛСА дозволяє отримати не тільки список публікацій, де зустрічається слово "інформація", але і без цього слова з найбільш близькими за змістом [80]. У табл. 3.5 представлений результат аналізу за такими ключовими словами: "марковский", "проект", "інформація". Ліва колонка відображає рівень схожості назви публікації з ключовими словами. Схожість

визначається відстанню між ключовим словом і документом в просторі, побудованому за допомогою ЛСА (сингулярне розкладання матриці). Для двовимірного простору розрахунок схожості (у%) може бути наступним:

$$S = \left(1 - \sqrt{\left(x_d - x_t\right)^2 + \left(y_d - y_t\right)^2}\right) \cdot 100, \tag{3.1}$$

де x_d та y_d – координати документа;

x_t та y_t – координати терма.

Таблиця 3.5 – Результат латентно семантичного аналізу за заданими ключовими словами

%	Наукометрична база	Назва публикації
88.75	Google academy	Матричная диаграмма и «сильная связность» индикаторов ценности в проектах
80.40	Google academy	Сетевые процессы управления проектами в контексте отображения состояний проекта
76.52	Base-search	Разработка марковских моделей изменений состояния пациентов в проектах предоставления медицинских услуг
69.47	Base-search	Трансформация когнитивных карт в модели марковских процессов для проектов создания программного обеспечения
55.57	Google academy	Управление знаниями в IT-проектах
51.27	Base-search	Составляющие поведенческой компетенции участника команды проекта на основе компетентностного подхода
47.99	Base-search	Анализ структурной модели компетенций по управлению проектами национального стандарта Украины
…		
-0.37	Base-search	К вопросу о патоморфологических исследованиях нейроэндокринной системы при завершенных суицидах

За цими результатами можна припустити, що публікації з високим рівнем схожості належать нашому авторові, а з низьким – іншим авторам.

Ще одним варіантом застосування латентно семантичного аналізу є розбиття документів на деякі групи, пов'язані за змістом. Наприклад, ми не можемо виділити ключові слова для напрямки наукової діяльності. За допомогою латентно семантичного аналізу можна проаналізувати список назв публікацій і розбити їх, припустимо, на 3 частини і надати для кожної групи набір ключових слів (табл. 3.6).

Таблиця 3.6 – Ключові слова, що відповідають смисловим групам

№ группи	Ключові слова
1	печен, неалкогольн, жиров, болезн
2	систем, исследован, суицид, патоморфологическ
3	проект, процесс, управлен, состоян

З цього списку ключових слів можна встановити, що третя група відноситься до тематики нашого автора. Виконавши пошук за ключовими словами цієї групи, можна отримати основну частину статей певного автора, без публікацій медичної тематики. Точніше, більш високий рівень схожості буде отримано для публікацій нашого автора [79]. Можна відкидати публікації, рівень схожості яких не перевищує наперед заданий поріг.

Ключові слова, запропоновані латентно семантичним аналізом, можна зберегти і наступного разу використовувати їх при новому пошуку публікацій цього учасника. Таким чином, можна створити навчальну систему в напівручному режимі і використовувати в автоматичному.

Результат латентно семантичного аналізу, звичайно ж, може мати похибку. Це добре помітно, коли ключові слова можна віднести до різних предметних областей наукової діяльності. Наприклад, слово "проект" може використовуватися в будь-якій сфері: навчальний проект, медичний проект, управління проектами і ін. При цьому в документах з малою кількістю слів, ключові слова можуть мати більшу вагомість.

Використання ЛСА дозволяє в деякій мірі автоматизувати і полегшити аналіз документів. З його допомогою можна виділяти схожі за змістом документи, класифікувати документи, а також отримувати ключові слова, що належать до різних смисловим групам.

На практиці це дозволяє вирішити проблему визначення достовірності авторства публікацій. Незважаючи на те, що в проекті по вилученню публікацій з наукометричних баз даних документи мають назви з декількох слів, застосування ЛСА дозволяє отримати позитивні результати.

3.5 Модель латентного розміщення Діріхле

Модель латентного розміщення Дирихле (latent Dirichlet allocation, LDA) – передбачає, що кожне слово в документі породжене деякою латентної темою, при цьому в явному вигляді моделюється розподіл слів у кожній темі, а також апріорне розподіл тем в документі. Теми всіх слів в документі передбачаються незалежними. У LDA, як і в PLSA, документ може відповідати не одній теми. Але LDA задає модель породження, як слів, так і документів, тому з'являється додаткова можливість оцінювати ймовірності документів поза текстової колекції за допомогою алгоритму варіаційного виведення і семпліювання по Гіббсу.

На відміну від PLSA, в LDA число параметрів не збільшується зі зростанням числа документів в колекції. Численні розширення моделі LDA усувають деякі її обмеження і покращують продуктивність для конкретних завдань (рис. 3.7). LDA безумовно лідирує серед імовірнісних тематичних моделей завдяки численним узагальнень і додатків до аналізу колекцій текстових документів:

– автор-тематична модель (author-topic model), яка розширює LDA для спільного опису документів і авторів;

– прихована тематична модель гіпертексту (latent topic hypertext model, LTHM) описує закон породження посилань в корпусі гіпертекстів;

– композитна модель HMM–LDA будується як спільний опис синтаксису і семантики тексту.

– прихована марківська модель (HMM) описує локальні закономірності між сусідніми словами, тоді як модель LDA дає глобальний тематичний опис документа в цілому.

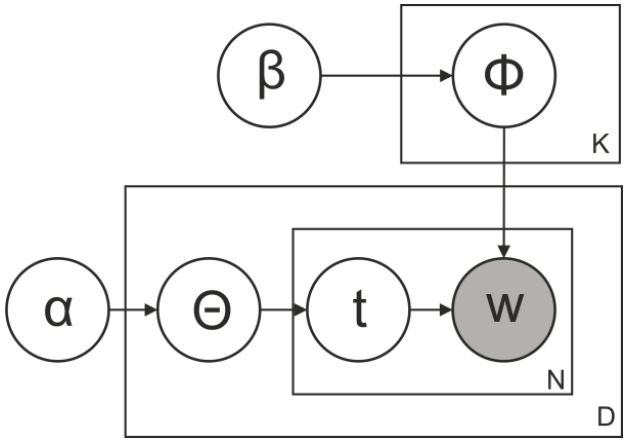


Рисунок 3.7 – Графічна імовірнісна модель LDA:

w – слово (змінна, що спостерігається);

t – тема (прихована змінна);

D – колекція документів;

N – довжина документа в словах;

K – число тем в колекції;

θ – розподіл тем в документі;

φ – розподіл слів в темі;

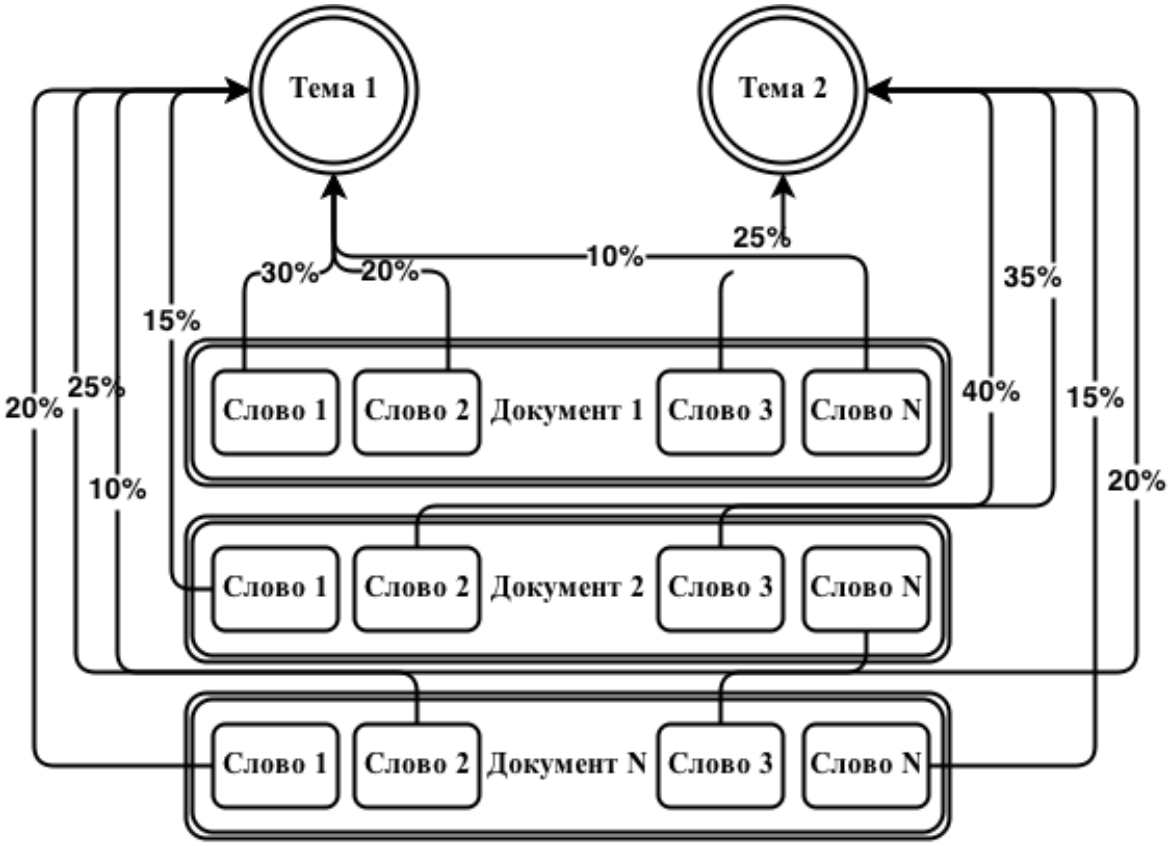
α – апріорний розподіл Діріхле на параметрі θ;

β – апріорний розподіл Діріхле на параметрі φ.

Виходячи з аналізу існуючих імовірнісних тематичних моделей, модель латентного розміщення Дирихле (LDA) є слушним кандидатом на використання в проекті по вилученню публікацій з наукометричних баз даних. Після вилучення публікацій з наукометричних баз даних ми отримуємо список їхніх назв. Завданням LDA є автоматичне визначення тематик, які містять ці назви. Застосування LDA дасть нам модель, показану на рис. 3.6. Показані розподіли слів в двох темах. Виходячи з цього, можна судити, що документ 1 більше відноситься

ся до першої темі, ніж до другої (30% + 20% + 10% проти 25%), а також отримати список найбільш підходящих до теми слів.

Для ідентифікації параметрів моделі LDA по колекції документів можна застосувати семпліювання по Гіббсу – це алгоритм для генерації вибірки спільного розподілу множини випадкових величин.



Малюнок 3.8 – Модель LDA з виділенням двох тем.

Цей алгоритм використовується для оцінки спільного розподілу і для обчислення інтегралів методом Монте-Карло. Припустимо, що слід визначити K тем в наборі документів, тоді алгоритм семпліювання за Гіббсом можна описати так:

1. Для кожного слова з кожного документу привласнити випадковим чином одну тему (t) з *K* можливих;

2. Для кожного слова з кожного документа обчислити:

– p(t|d) – пропорція слів в документі d, які присвоєні темі t;

– p(w|t) – пропорція слова w у всіх документах, присвоєного до теми t;

– привласнити слову w нову тему t з імовірністю p(t|d) * p(w|t).

3. Повторити другий пункт кілька разів (кількість ітерацій також є вхідним параметром).

Метод LDA заснований на наступній ймовірнісної моделі:

$$p(d, w) = \sum_{t \in T} p(d) \cdot p(w \mid t) \cdot p(t \mid d), \quad (3.2)$$

де d – документ;

t – тема;

w – слово;

T – множинао тем;

p(d) – апріорний розподіл на множині документів;

p(w|t) – умовний розподіл слова w в темі t;

p(t|d) – умовний розподіл теми t в документі d.

Для демонстрації результатів аналізу, візьмемо сукупність публікацій, отриманих по параметру пошуку «Яковенко В. Д.» і виконаємо порівняння з такими ключовими словами: «система», «автоматизоване». Процедура виконана також з використанням LSA двома повтореннями. Результати показані в табл. 3.7.

Таблиця 3.7 - Результат порівняння публікацій з ключовими словами

LSA, %		LDA, %		Публикація
1	2	1	2	
82	82	100	100	Прогнозирование состояния системы управления качеством деятельности учебного заведения
0	0	52	53	К вопросу о причинно-следственных взаимосвязях в патогенезе хронического тонзиллита, как инфекционно-аллергического процесса
0	0	52	53	Некоторые закономерности соотношения дефицита барьерной функции миндалин и системного иммунитета при хроническом тонзиллите
99	99	41	54	Прогнозування стану системи керування якістю навчального закладу
87	87	96	36	Комп'ютерна реалізація системи автоматизованого управління навчальним процесом
88	88	96	45	Формалізація вимог до системи автоматизованого управління навчальним закладом

З таблиці слідує, що результати LSA і LDA в деякому роді схожі, але через те, що LDA використовує випадкові величини, результати можуть відрізнятися на одних і тих же вхідних документах. Також, через малу кількість документів, LDA показує досить великий відсоток схожості для документів, що не відповідають заданим ключовим словам.

Таким чином, можна зробити висновок, що для проекту щодо вилучення публікацій з наукометричних баз даних, латентно семантичний аналіз підходить краще, ніж імовірнісна модель.

Через невеликий обсяг, як публікацій, так і їх вмісту (назв публікацій в нашому випадку), імовірнісна модель латентного розміщення Дирихле показує найгірші результати. З огляду на те, що одним з недоліків LSA є зниження швидкості обчислення при збільшенні обсягу даних, для цього проекту його не слід застосовувати.

Латентне розміщення Діріхле є базовою ймовірнісної тематичної моделлю і лідирує серед інших завдяки численним узагальнень і додатків до аналізу колекцій текстових документів. Базові ймовірнісні тематичні моделі дозволяють виявляти приховану тематику документів на основі моделі документа як мішка слів. У них також передбачається існування прихованих взаємозв'язків між різними об'єктами, які можуть проявлятися в структурі слововживання. Семантична близькість різних об'єктів може оцінюватися шляхом порівняння їх тематичних векторів. Але, застосувавши латентне розміщення Діріхле до проекту щодо вилучення публікацій з наукометричних баз даних, помічено, що результати гірші, ніж ті, які визначаються з використанням латентно семантичного аналізу. Тому, не дивлячись на недоліки ЛСА, використання його в цьому проекті є виправданим.

3.6 Висновки

Моделі з прихованими (латентними) змінними є особливо ефективними для виявлення прихованих структур в текстових колекціях. Графічні моделі з прихованим тематичним шаром останнім часом успішно застосовувалися в задачах пошуку прихованих закономірностей у даних. Автоматичне виділення тематики текстів застосовувалося для розбиття текстів по групах на основі семантичної близькості змісту. Одним з перших був запропонований імовірнісний латентний семантичний аналіз, слідом за ним – метод латентного розміщення Дирихле і його численні узагальнення. Застосування тематичних моделей дозволяє отримати відповідь на цілий ряд нетривіальних питань, таких як: виявляти сенс або тематику документів за їх змістом, здійснювати класифікацію документів на основі цих прихованих тематичних закономірностей, виявляти свої інтереси авторів та інші.

Латентно семантичний аналіз являє модель дворежимного факторного аналізу, яка заснована на сингулярному розкладанні. Сингулярне розкладання представляє терміни і документи у вигляді векторів в просторі обраній розмірності, а скалярний добуток між точками простору – їх схожість. Латентно-семантичний аналіз добре справляється з проблемою синонімії, але частково з проблемою полісемії, тому, що кожне слово визначається однією точкою в просторі.

Застосування латентно семантичного аналізу в даному дослідженні дозволяє вирішити проблеми ідентифікації публікацій певного автора, а також виділити найбільш вагомі ключові слова, які зустрічаються в назвах його публікацій. Назви публікацій містять відносно мала кількість слів, щоб латентно семантичний аналіз працював з мінімально похибкою. Не дивлячись на це, його застосування все ж показує позитивний результат.

4 РОЗРОБКИ ПРОГРАМНОГО ПРОДУКТУ ДЛЯ ВИТЯГАННЯ І ОБРОБКИ ІНФОРМАЦІЇ З НАУКОМЕТРИЧНИХ БАЗ ДАНИХ

4.1 Основні вимоги до програмного продукту

Завданням даного програмного продукту є надати список публікацій здобувача, які індексуються в міжнародних наукометричних базах даних.

Однією з перших стадій розробки програмного проекту є збір інформації, аналіз, специфікація, і перевірка вимог до програмного забезпечення. Програмні вимоги – властивості програмного забезпечення, які повинні бути належним чином представлені в ньому для вирішення конкретних практичних завдань. Досвід індустрії інформаційних технологій однозначно показує, що питання, пов'язані з управлінням вимогами, надають критично-важливий вплив на програмні проекти, певною мірою і на сам факт можливості успішного завершення проектів.

Вимогами до даного проекту є:

– витяг інформації з Веб сторінок;

– критерієм інформації є ПІБ автора;

– робота з найбільш поширеними наукометричними базами даних: Scopus, Web of Science;

– обробка результатів з метою визначення нерелевантної інформації;

– надання інформації користувачеві.

Множина чинників в слабо структурованих системах створення програмних проектів утворює складну «павутину» зв'язків і станів, що змінюються в часі. Розвиток програмних проектів у такий багатофакторній системі, як правило, вдається представити тільки в формі якісних моделей [1]. Разом з тим трансформація якісних когнітивних моделей в марковські моделі дозволить перейти до кількісних оцінок ходу і результатів проектів [102].

4.2 Трансформація когнітивних карт в моделі марківських процесів для проектів створення програмного забезпечення

Управління розвитком проектно-керованих або проектно-орієнтованих систем припускає наявність підсистем аналізу, підготовки та прийняття ефективних рішень [104]. Створення таких інформаційних технологій для управління проектами / програмами / портфелями проектів неможливо без розробки моделей, методів і механізмів взаємодії команди проекту.

Вихідним поняттям когнітивного моделювання проектів і складних процесів є когнітивна карта, яка являє собою орієнтований зважений граф, в якому [105]:

- вершини відповідають базисним факторам (станам) проекту, які можуть бути ідентифіковані з використанням технології data mining, для обґрунтованого відкидання надлишкових чинників, які слабо впливають на інші фактори проекту;
- безпосередні зв'язки факторів відображають причинно-наслідкові ланцюги, по яких поширюються впливи деякого фактора на інші чинники та вважається, що фактори, які входять до умови «якщо ..., тоді ...», впливають на чинники слідства всього ланцюга; за таких умов вплив може бути або підсилюючим (позитивним), або гальмуючим (негативним), або змінного знаку у залежності від множини додаткових умов проекту.

Когнітивна карта віддзеркалює лише структуру зв'язків між факторами. У ній не відбивається сутність впливу, а також динаміка впливів у разі зміни ситуації або зміни в часі самих чинників. Відображення цих особливостей, відображеної в когнітивної карті, можливо на наступному рівні структуризації інформації в когнітивної моделі [106]. На цьому рівні кожна комунікація між чинниками може бути розкрита у формі відповідного рівняння, до якого можуть бути включені кількісні або вимірювані параметри, а також і якісні або нечіткі висловлювання. За таких умов кількісні параметри природним чином відображаються у рівняннях через чисельні величини. Кожна якісна

змінна відображається у формі сукупності лінгвістичних висловлювань, які відображають різні значення цієї змінної (наприклад, функціональність програмного продукту може бути «низькою», «задовільною», «вище вимог технічного завдання на проект»), а кожної лінгвістичної змінної відповідає певний числовий еквівалент за шкалою [0,1], наприклад, з використанням функції бажаності Харрінгтона [107]. У разі накопичення нових знань стає можливим більш детально розкривати характер зв'язків між факторами [108].

Важливим етапом когнітивного моделювання є побудова когнітивної карти, яка являє подобу орієнтованого графа з вершинами, що відповідають базисним факторам (станам) проекту, і дугами, що відображають наслідкові зв'язки факторів [109 – 110]. При цьому знак «+» означає позитивний зв'язок, а «–» відповідає негативному зв'язку [105].

Розглянемо побудову когнітивної карти на прикладі управління проектом розробки програмного забезпечення (ПЗ).

Найбільш поширеним підходом до розробки ПЗ у даний час є версіонування, при якому послідовно виконуються етапи розробки та налагодження програмного коду, а оцінка результату зводиться до формули «як вийде». Такий підхід, як правило, забезпечує розробку ПЗ при прийнятних витратах і якості, але цей процес включає в себе безліч випадкових помилок і проб, є «знанням команди» і тримається на конкретних виконавців [111].

Згідно SWEBOK 2004 розробка ПЗ включає в себе використання 10 основних галузей знань [112]:

1. Software requirements – програмні вимоги.
2. Software design – дизайн (архітектура).
3. Software construction – конструювання ПЗ.
4. Software testing – тестування.
5. Software maintenance – підтримка ПЗ.
6. Software configuration management – конфігураційне управління.
7. Software engineering management – управління програмною інженерією.

8. Software engineering process – процеси програмної інженерії.
9. Software engineering tools and methods – інструменти й засоби.
10. Software quality – якість ПЗ.

Когнітивна карта розробки ПЗ включає 10 вершин, що відповідають основним областям знань, компетенціям та зв'язкам між цими вершинами (рис. 4.1).

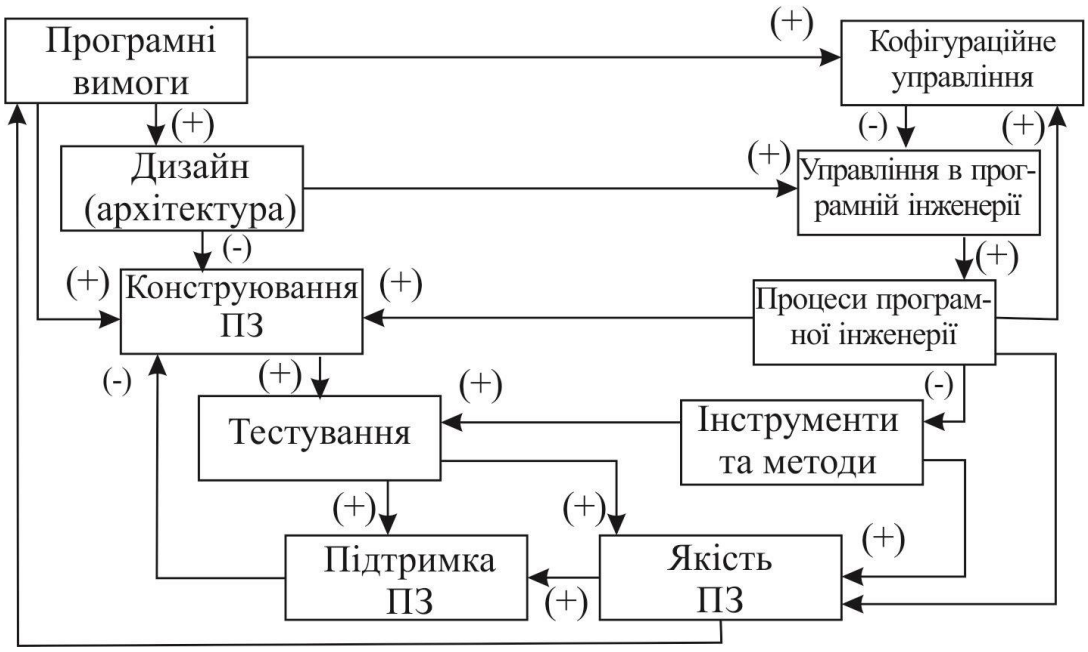


Рисунок 4.1 – Когнітивна карта розробки ПЗ

Фактично така когнітивна карта розробки ПЗ відображає стани системи і переходи між цими станами. Якщо прийняти, що сума ймовірностей всіх станів дорівнює одиниці, а також те, що переходи з кожного стану до іншого є несумісними подіями, то такий граф може бути трансформованим у ланцюг Маркова з дискретними визначеними станами і дискретним часом [113 – 115].

Для цього доповнимо орієнтований граф, що відображає когнітивні особливості проектів розробки ПЗ, зв'язками затримки в кожному з 10 процесів (станів) і отримаємо марківський ланцюг.

Зазначена трансформація когнітивної карти в марківский ланцюг дозволяє перейти від якісних оцінок протікання проектів до кількісних характеристик.

При цьому кількісні оцінки являють собою не тільки багатовекторну картину стану проектів, але і мають властивість прогнозування.

Матриця умовних перехідних ймовірностей для цього марківського ланцюга (рис. 4.1) матиме вигляд:

$||\pi_{ij}|| =$

$\pi_{1.1}$	$\pi_{1.2}$	0	0	0	0	0	0	0	$\pi_{1.10}$
0	$\pi_{2.2}$	$\pi_{1.3}$	0	0	0	0	0	$\pi_{1.9}$	0
0	0	$\pi_{3.3}$	$\pi_{3.4}$	0	0	0	0	0	0
0	0	0	$\pi_{4.4}$	$\pi_{4.5}$	$\pi_{4.6}$	0	0	0	0
0	0	$\pi_{5.3}$	0	$\pi_{5.5}$	0	0	0	0	0
$\pi_{6.1}$	0	0	0	$\pi_{6.5}$	$\pi_{6.6}$	0	0	0	0
0	0	0	$\pi_{7.4}$	0	0	$\pi_{7.7}$	0	0	0
0	0	$\pi_{8.3}$	0	0	$\pi_{8.6}$	$\pi_{8.7}$	$\pi_{8.8}$	0	0
0	0	0	0	0	0	0	$\pi_{9.8}$	$\pi_{9.9}$	0
0	0	0	0	0	0	0	0	$\pi_{10.9}$	$\pi_{10.10}$

Значення перехідних ймовірностей $\pi_{i.j}$ винайдемо за допомогою експертних методів:

$||\pi_{ij}|| =$

0,4	0,4								0,2
	0,5	0,3						0,2	
	0	0,2	0,8						
			0,4	0,4	0,2				
		0,6		0,4					
0,1		0		0,2	0,7				
			0,3			0,7			
		0,1			0,15	0,3	0,45		
							0,2	0,8	
								0,6	0,4

На підставі матриці перехідних ймовірностей, за умови, що початковий стан системи відомо, знайдемо всі ймовірності станів $p_1(k)$, $p_2(k)$,..., $p_{10}(k)$ після будь-якого k-го кроку [102]:

$$p_i(k+1) = \sum_{j=1}^{n} [p_j(k) \cdot \pi_{ji}]\Big|_{n=6}; \quad i = 1, 2, \cdots, n$$

На рис. 4.2 наведено результати моделювання станів системи для вихідної матриці умовних перехідних ймовірностей.

Для даного рівня компетентності та організованості команди проекту, відповідних сукупності значень перехідних ймовірностей, визначених експертним методом, можна зробити такі висновки. Найбільша ймовірність стану для умов $\pi_{3.4} = 0{,}4$ та $\pi_{3.3} = 0{,}6$ після 10-го кроку відповідає процесу 3 – «Конструювання ПЗ» (рис. 4.2) Далі найбільш ресурсоємними є процеси 4 і 5. «Управління якістю ПЗ» також можна віднести до найбільш важливих процесів – крива 6.

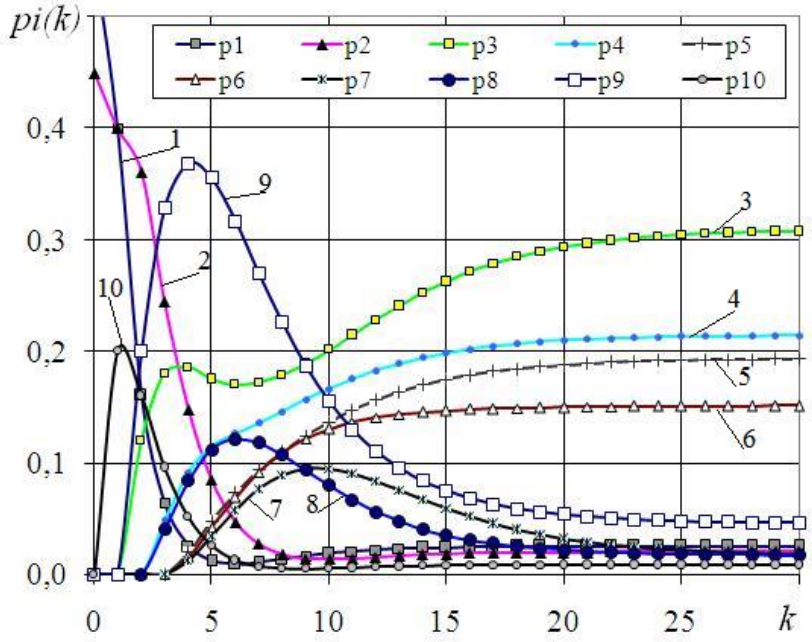


Рисунок 4.2 – Зміна ймовірності станів процесів для умов $\boldsymbol{\pi_{3.4}} = 0{,}4$ та $\boldsymbol{\pi_{3.3}} = 0{,}6$: 1 – програмні вимоги; 2 – дизайн (архітектура); 3 – конструювання ПЗ; 4 – тестування; 5 – підтримка ПЗ; 6 – конфігураційне управління; 7 – управління програмною інженерією; 8 – процеси програмної інженерії; 9 – інструменти й засоби; 10 – якість ПЗ.

Для умов $\pi_{3.4} = 0{,}4$ и $\pi_{3.3} = 0{,}7$ після 10-го кроку картина результативності проекту істотно змінюється – основними витратними процесами за часом стають процеси 4 (Тестування) та 5 (Підтримка ПЗ). На третю і четверту позицію

відповідно до затрат часу переміщаються, відповідно, процеси 3 (Конструювання ПЗ) та 6 (Управління якістю ПЗ) (рис. 4.3).

Отримані ймовірності станів дозволяють прогнозувати та оцінювати результативність виконання проектів. Зазначимо, що по мірі виконання проекту ступінь ресурсоємності окремих процесів зміняться.

Основною причиною більшості провалів програмних проектів є саме застосування неадекватних методів управління його розробкою.

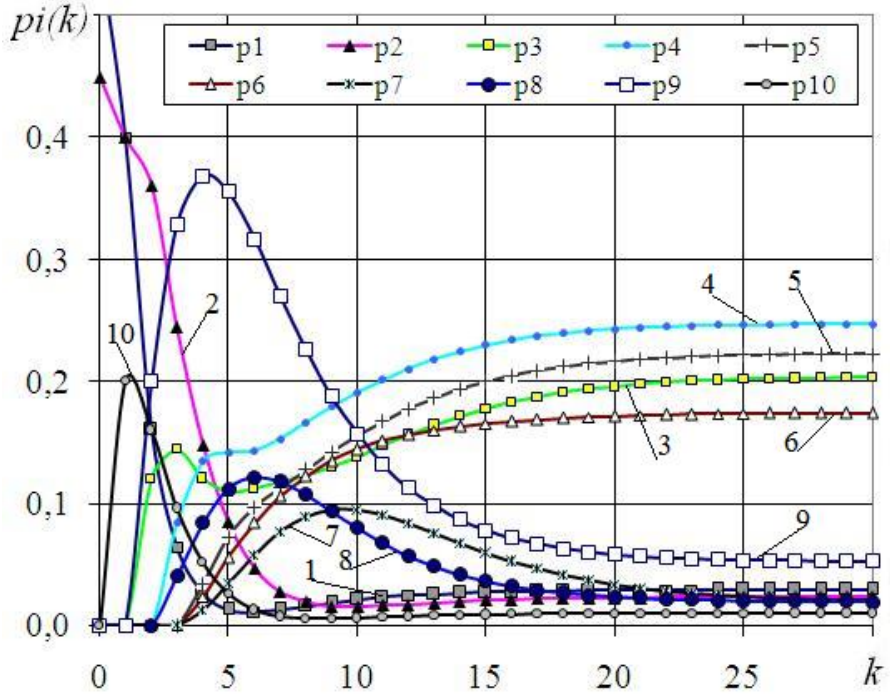


Рисунок 4.3 – Зміна ймовірності станів процесів для умов $\pi_{3.4}$ = 0,3 та $\pi_{3.3}$ = 0,7: 1 – програмні вимоги; 2 – дизайн (архітектура); 3 – конструювання ПЗ; 4 – тестування; 5 – підтримка ПЗ; 6 – конфігураційне управління; 7 – управління програмною інженерією; 8 – процеси програмної інженерії; 9 – інструменти й засоби; 10 – якість ПЗ.

Класичні підходи до управління складними системами, що пов’язані зі створенням програмних продуктів, не «працюють» у випадках, якщо структурні або параметричні характеристики об'єкта керування не відомі або суттєво

змінюються в часі [116]. Ці підходи також не допоможуть, якщо поточні властивості об'єкта не дозволяють йому розвиватися з необхідними характеристиками. Якщо команда проекту не може забезпечити необхідну ефективність і тому постійно працює в режимі авралу, то це призводить не до зростання продуктивності, а до відходу професіоналів з проекту [116].

Застосування когнітивних карт з подальшим їх відображенням за допомогою марківських ланцюгів дозволяє кількісно представити хід проектних процесів, що є істотною умовою успішності виконання проектів. Область застосування запропонованого методу трансформації когнітивних карт в марковские моделі може бути істотно розширена за рахунок застосування в навчальному процесі для компетентнісного навчання при підготовці фахівців.

4.3 Програмний проект

Процес визначення архітектури, компонентів, інтерфейсів та інших характеристик системи або її компонентів називається проектуванням. Результат процесу проектування – дизайн. Проектування є інженерна діяльність, в якій належним чином аналізуються вимоги для створення опису внутрішньої структури ПО і є основою для його конструювання. Програмний дизайн (як результат діяльності з проектування) повинен описувати архітектуру програмного забезпечення, тобто представляти декомпозицію програмної системи у вигляді організованої структури компонент і інтерфейсів між компонентами. Найважливішою характеристикою готовності дизайну є той рівень деталізації компонентів, який дозволяє зайнятися їх конструюванням. Проектування програмних систем можна розглядати як діяльність, результат якої складається з двох складових частин:

–Архітектурний або високорівнева дизайн – опис високорівневою структури і організації компонентів системи;

– Деталізований дизайн – описує кожен компонент в тому обсязі, який необхідний для конструювання.

Розділяють такі види дизайну:

– D-дизайн – декомпозиція структури програмного забезпечення у вигляді набору фрагментів або компонент;

– FP-дизайн – сімейство архітектурних уявлень, що базуються на шаблонах;

– I-дизайн – створення високо-рівневої концепції, бачення того, що з себе представлятиме програмна система; даний вид дизайну є результатом процесу аналізу вимог і їх трансформації в підходи до реалізації.

Проектування програмного забезпечення в розумінні програмної інженерії має на увазі D- і FP-дизайн. I-дизайн більшою мірою відноситься до роботи з програмними вимогами.

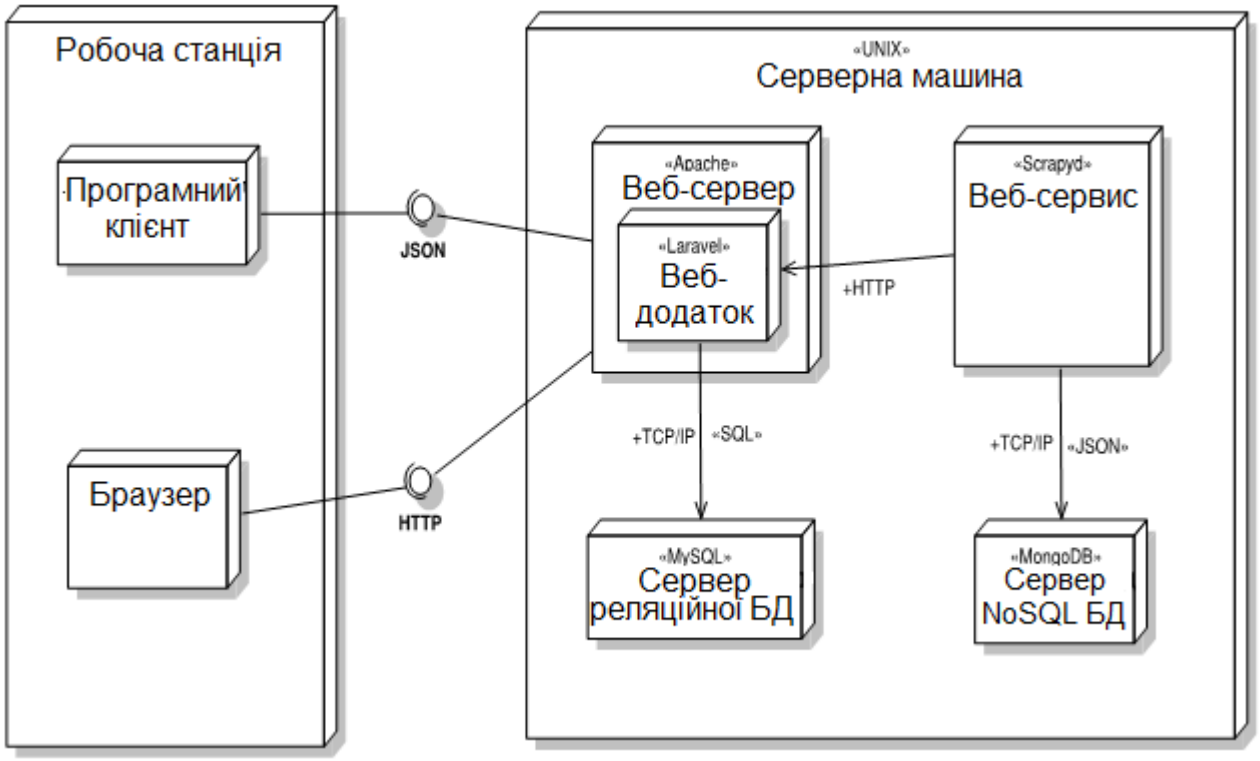


Рисунок 4.4 – Архітектура програмного проекту з вилучення публікацій

Декомпозиція структури програмного проекту у вигляді набору компонент представлена на рис 4.4.

Проект являє собою програмний комплекс з декількох додатків, взаємодія яких надає сервіс пошуку та вилучення публікацій зазначеного автора (табл. 4.1).

Таблиця 4.1 – Використовувані засоби і технології

Категорія	Значення	Де використовується
Мови програмування	PHP, Javascript	Веб додаток
	Python	Веб сервіс
Фреймворки, бібліотеки	Laravel	Веб додаток
	Guzzle	Веб додаток
	Scrapy	Веб сервіс
	Selenium WebDriver	Веб сервіс
	Gensim	Веб сервіс
	NLTK	Веб сервіс
Зовнішні додатки	Apache web server	Веб додаток
	PhantomJS	Веб сервіс
	MySQL server	Веб додаток
	MongoDB server	Веб сервіс
	Scrapyd	Веб сервіс

У даному проекті використовуються кілька мов програмування, різні бібліотеки і додатки, які відображені в табл. 4.1. Колонка «Де використовується» показує який з двох основних компонентів використовує цю технологію.

Основними компонентами системи є:

– Веб додаток smd;

– Веб сервіс scrapyd;

Додаткові компоненти, з якими працюють основні це:

– сервер реляційної БД MySQL;

– сервер NoSQL БД MongoDB.

Веб додаток SMD являє собою графічний інтерфейс користувача, а також надає програмний інтерфейс для використання пошуку публікацій іншими додатками. Веб сервіс Scrapyd представляє сервіс по вилученню структурованих даних з НМБД, а також управляє запуском відповідних програм-павуків окремої для кожної НМБД. Таким чином, функціонал програмної системи розділений на окремі модулі – додатки, які працюють незалежно один від одного. Веб додаток SMD використовує сервіс Scrapyd під час для пошуку публікацій за запитом користувача. Ці програми спілкуються між собою по HTTP протоколу в JSON форматі.

Веб додаток SMD використовує реляційну базу даних (MySQL) в якості сховища даних, таких як інформація про користувачів, список підтримуваних НМБД, історія результатів пошуку публікацій та ін. Веб сервіс Scrapyd використовує документо-орієнтовану базу даних (NoSQL) для тимчасового зберігання результатів пошуку на зовнішньому диску, таким чином, не збільшуючи об'єм використання оперативної пам'яті при витяганні великої кількості публікацій. Доступ до баз даних надають окремі додатки – СУБД, з якими програми працюють по протоколу TCP/IP. Робота з додатком виконується за допомогою веб браузера. Також є програмний доступ до інтерфейсу у форматі JSON.

Основними варіантами використання програми, які показані на рис. 4.5, є:

– реєстрація користувачів в системі – створення облікового запису користувача для можливості прив'язки знайдених публікацій до користувача;

– пошук публікацій – основний варіант використання. З одного боку користувач запускає пошук по заданих параметрах, з іншого боку сервіс пошуку (scrapyd), який керує цим процесом. Основні етапи пошуку публікацій це вилучення інформації, її аналіз (включаючи латентно-семантичний) і збереження результатів;

Рисунок 4.5 – Варіанти використання проекту по вилученню публикацій

– історія пошуку публікацій – навігація по історії виконаних пошукових запитів;

– перегляд результатів пошуку складається з двох варіантів використання;

– прив'язка публікацій до користувача і відображення статистики по знайденим публікаціям або публікаціям прив'язаних до користувача.

Характеристика використаних засобів і технологій:

PHP (PHP: Hypertext Preprocessor) – скриптова мова програмування загального призначення, інтенсивно застосовується для розробки веб-додатків.

JavaScript – прототипно-орієнтований сценарний мову програмування. JavaScript зазвичай використовується як вбудований мова для програмного доступу до об'єктів додатків. Найбільш широке застосування знаходить в браузерах як мова сценаріїв для додання інтерактивності веб-сторінок.

Python – високорівнева мова програмування загального призначення, орієнтований на підвищення продуктивності розробника і читання коду.

Laravel – безкоштовний веб-фреймворк з відкритим кодом, призначений для розробки з використанням архітектурної моделі MVC (Model View Controller – модель–уявлення–контролер).

Guzzle – бібліотека для PHP за допомогою якої легко слати HTTP запити і неважко інтегрувати додаток з веб сервісами.

Scrapy це фреймворк для обходу веб-сайтів і вилучення структурованих даних, які можуть бути використані для широкого додатків.

Selenium – це інструмент для тестування Web-додатків. Selenium WebDriver API використовується для доступу до браузеру.

Gensim є бібліотекою мовою програмування Python і призначена для автоматичного вилучення семантичних тим з документів. Алгоритми в gensim: латентного семантичний аналіз, латентний розподілу Діріхле.

NLTK (Natural Language Toolkit) – набір бібліотек і програм для символічної і статистичної обробки природної мови на мову програмування Python.

Apache HTTP-сервер – вільний веб-сервер.

PhantomJS – скриптова браузер без графічного інтерфейсу, який використовується для автоматизації взаємодії з веб-сторінками.

MySQL – вільна реляційна система управління базами даних.

MongoDB – документо-орієнтована система управління базами даних (СУБД) з відкритим вихідним кодом, що не вимагає опису схеми таблиць.

Scrapyd являє собою додаток для розгортання і запуску SCRAPY павуків. Це дозволяє розгортати ваші проекти і контролювати своїх павуків за допомогою JSON API.

Даний програмний продукт розроблено як один з інструментів інформаційного забезпечення моніторингу публікаційної активності науковців (рис. 4.6).

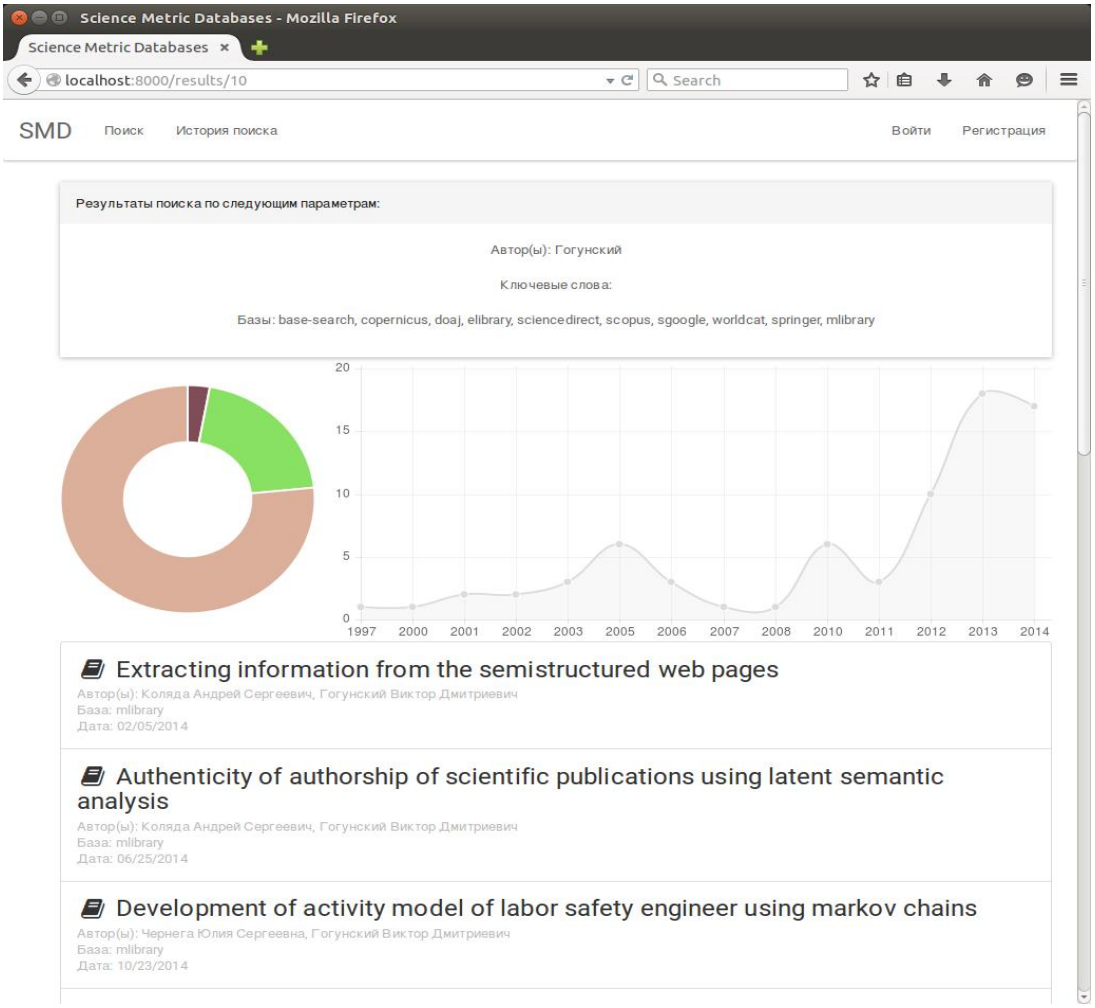


Рисунок 4.6 – Екраний інтерфейс системи Science Metric Databases

Система Science Metric Databases надає метадані публікацій, які індексуються в міжнародних наукометричних базах даних. Основними особливістю програмного продукту є: можливість вилучення інформації з неструктурованих даних (Веб-сторінок) і обробка цієї інформації з метою визначення нерелевантної інформації та фільтрації її.

Даний програмний продукт складається з Веб-додатку і Веб-сервісу, які взаємодіють між собою. Веб-сервіс призначений для пошуку і вилучення публікацій, а Веб-додаток надає графічний інтерфейс користувача, який відображає знайдені публікації і містить інтерфейс ініціалізації пошуку. Веб-додаток також надає програмний інтерфейс для можливого його автоматизованого використання.

4.4 Висновки

Наукометрія – це наука, завданням якої є вимір і аналіз наукових досліджень (статистика, кількість публікацій, індекс цитування та інші). Розподіл вчених за кількістю публікацій дозволяє не тільки виявити продуктивність, але і визначити ранг вченого, і, отже, його значимість. Це допомагає, наприклад, обґрунтувати включення робіт даного дослідника в список літератури свого дисертаційного дослідження. Розподіл публікацій за науковими напрямами для різних країн дає можливість отримати уявлення про відносний рівень розвиненості окремих галузей науки в країнах може бути використано при виробленні рішення про вивчення публікацій тієї чи іншої країни в рамках своєї дослідницької роботи. Також активність публікації наукових співробітників є одним з основних факторів, який враховується при визначенні світових рейтингів вищих навчальних закладів.

З розвитком інформаційних технологій з'явилися спеціалізовані засоби для автоматизації наукометричної діяльності, які називають наукометричними базами даних. Найбільш відомі і великі з них є Scopus і Web of Science. Також серед некомерційних наукометричних баз даних можна виділити Copernicus,

BASE, DOAJ, Science Index, WorldCat, MLibrary. Інформація про публікації в цих базах зберігається у вигляді метаданих – структурованих даних, що представляють характеристики публікації для цілей їх ідентифікації, пошуку, оцінки або управління. Використовуючи ці метадані можна витягувати публікації певного наукового співробітника.

Даний програмний продукт розробляється як один з інструментів моніторингу активністю публікацій наукових співробітників. Завданням його є надати метадані публікацій здобувача, які індексуються в міжнародних наукометричних базах даних. Основними вимогою до програмного продукту є: можливість отримання інформації з неструктурованих даних (веб сторінок) і обробка цієї інформації з метою визначення нерелевантной інформації та фільтрації її.

Виконано дослідження, щодо технології створення програмних продуктів. Показано, що застосування когнітивних карт з подальшим їх відображенням за допомогою марківських ланцюгів дозволяє кількісно представити хід проектних процесів, що є істотною умовою успішності виконання проектів. Область застосування запропонованого методу трансформації когнітивних карт в марковські моделі може бути істотно розширена у разі застосування в навчальному процесі для компетентнісного навчання при підготовці фахівців

Програмний продукт складається з веб додатки і веб сервісу, які взаємодіють між собою. Веб сервіс займається пошуком і витяганням публікацій, а веб додаток надає графічний інтерфейс користувача, який відображає знайдені публікації і містить інтерфейс ініціалізації пошуку. Веб додаток також надає програмний інтерфейс для можливого автоматизованого використання його.

Для реалізації процесів вилучення інформації з Веб сторінок використовується фреймворк Scrapy, який дозволяє швидко створити програму павука. Для обробки динамічних сторінок використовується безголовий браузер PhantomJS. Результати витягнутих даних тимчасово зберігаються в NoSQL базу даних MongoDB, а потім перетворюються в реляційні дані і зберігаються в БД MySQL. За реалізацію веб додатки відповідає веб фреймворк Laravel, за допомогою якого за невеликий час можна створити стабільно працююче додаток.

ВИСНОВКИ

В досліджені вирішена актуальна науково-прикладна задача теоретичного обґрунтування моделей і методів аналізу контенту Веб сторінок з імітацією роботи користувачів для автоматизованого вилучення з наукометричних баз за допомогою розроблених програмних інструментів метаданих наукових статей.

Одним з напрямів діяльності МОН України щодо входження до світового наукового співтовариства є створення інструментів «вимірювання» активності публікацій вчених з подальшим формуванням інформаційно-аналітичної системи моніторингу публікацій вчених ВНЗ України. Для реалізації цих інструментів потрібен спосіб вилучення метаданих публікацій з наукометричних баз, що і є завданням даного дослідження.

Веб інтерфейс наукометричних баз даних є універсальним, а часто і єдиним, способом доступу до метаданих публікацій. Тому для розвязання завданнь вилучення метаданих публікацій використовується Веб скрапінг – процес аналзу інформації з Веб сторінок, який фокусується на перетворенні неструктурованих даних в мережі (наприклад, у форматі HTML) в структурований формат даних, який може бути проаналізований і збережений для подальшого використання. Веб-скрапінг використовує програми для обходу і завантаження Веб-сторінок по заданому критерію. На відміну від пошукових машин, сканується вузьке коло веб-сторінок, заданих початковими умовами і витягується тільки потрібна інформація. В результаті даного дослідження спроектована система вилучення інформації про наукові публікації по параметру пошуку «Автор». Використовуючи цю властивість, виконується пошук по найбільш відомим наукометричним базам даних, і витягуються метадані певної (даної системи) структури.

Реалізація завантаження Веб-сторінок, навігація по посиланнях і вилучення даних з веб ресурсів проводиться за допомогою Веб скрапінг фреймворка Scrapy. Використовуваний набір технологій і програмного забезпечення дозволяють створити програмний продукт по вилученню інформації з неоднорідних і

неформалізованих джерел (таких як наукометричних баз) і перетворення її в структурований вигляд з можливою подальшою обробкою. Ці дані можуть використовуватися аспірантами і здобувачами, наприклад, при підготовці до захисту дисертацій. Крім того пропонована система може бути корисна при оцінці діяльності ВНЗ.

Після вилучення метаданих публікацій виникає проблема їх обробки – визначення публікацій конкретного автора, відкинувши результати, наприклад, однофамільців. Для вирішення цієї проблеми підходить модель з прихованими (латентними) змінними, яка є особливо ефективними для виявлення прихованих структур в текстових наборах. Автоматичне виділення тематики текстів можна застосувати для розбиття текстів по групах на основі семантичної близькості змісту. У нашому випадку текстовим набором є назви публікацій. Латентно семантичний аналіз є моделлю, яка використовується в даному дослідженні. Застосування її дозволило вирішити проблеми ідентифікації публікацій певного автора, а також виділити найбільш вагомі ключові слова, які зустрічаються в назвах його публікацій. Не дивлячись на те, що назви публікацій містять відносно мала кількість слів, застосування латентно семантичного аналізу показує позитивний результат.

Таким чином, дане дослідження було сфокусовано на добуванні метаданих наукових публікацій з наукометричних баз даних та їх обробки їх з метою фільтрації нерелевантних результатів (наприклад, публікації однофамільців). Результати роботи програмного комплексу можуть бути використані далі для створення інструментів моніторингу активністю публікацій наукових співробітників.

Отримані наступні результати.

1 *Внесок у теоретичні основи інформаційних технологій:*

1.1. На основі аналізу опублікованих робіт та існуючих програмних продуктів і інструментів наукометричних баз встановлено, що пошук публікацій в наукометричних базах даних, як правило, здійснюється тільки в межах окремих

баз даних або репозитаріїв, що не дозволяє визначити інтегральну оцінку публікаційної активності науковців.

1.2. Виконана формалізація інформаційної технології для задач управління пошуком метаданих публікацій в наукометричних базах даних, що включає сучасну комп'ютерну систему накопичення, переробки і збереження інформації, що дозволяє розробити і впровадити Інтернет-технологію для побудови сервіс-орієнтованої системи інформаційного забезпечення кінцевих користувачів;.

1.3. Обґрунтована і розроблена інформаційно-пошукова система автоматизації вилучення метаданих публікацій з поширених наукометричних баз даних, яка включає програмні інструменти вилучення та аналізу контенту Веб сторінок, що дозволяє виконати інтегральну оцінку публікаційної активності авторів наукових публікацій;

1.4. Удосконалено метод Дірихле та модель латентно-семантичного аналізу, що містять ймовірнісні оцінки та інструментальні засоби класифікації і визначення достовірності інформації, що вилучається з контенту Веб сторінок, і засновані на аналізі прихованих змінних для виявлення зв'язків в наборі назв публікацій, що дозволяє достовірно ідентифікувати публікації конкретних авторів.

2 *Внесок в методи побудови інформаційно-пошукових систем:*

2.1. Запропонована концепція побудови інформаційно-пошукових систем і способів інформаційного забезпечення користувачів, яка базується на інформаційній технології вилучення та аналізу контенту Веб сторінок наукометричних баз даних, що дозволяє виконувати моніторинг інтегральної публікаційної активності, як окремих науковців, так і наукових колективів.

2.2. Розроблені програмні інструменти вилучення інформації з Веб сторінок, які конструюються динамічно на стороні користувача (клієнта), що дозволяє побудувати інформаційну технологію витягання контенту з елементами інтелектуальності в умовах невизначеності.

2.3. Розроблено програмний продукт, який реалізує інформаційну технологію пошуку публікацій науковців у найбільш відомих наукометричних базах

даних; програмний продукт може бути корисним, як навчальним закладами, так і окремим науковцям, яким потрібно знати які їх публікації індексуються певними наукометричними базами даних.

3 *Створення передумов для подальших досліджень:*

3.1. Результати дисертаційних досліджень можуть бути основою для розвитку інформаційних технологій щодо забезпечення інформаційних потреб окремих науковців зі створенням інформаційно-пошукових систем для більшого числа наукометричних баз даних.

3.2. Запропонована і розроблена інформаційна технологія, яка в роботі орієнтована на забезпечення особистих інформаційних потреб окремих науковців, може бути формалізована, як програмний додаток (APA), для включення в інші програмні комплекси для моніторингу публікаційної активності лабораторій, кафедр, університетів.

ПЕРЕЛІК ПОСИЛАНЬ

1. Бурков, В. Н. Параметры цитируемости научных публикаций в наукометрических базах данных / В. Н. Бурков, А. А. Белощицкий, В. Д. Гогунский // Управління розвитком складних систем. – 2013. – № 15. – С. 134 – 139. doi: doi.org\10.13140/RG.2.1.3092.8087
2. Оборский, Г.А. Наукометрические исследования публикационной активности как составляющая инновационного развития университета / Г.А. Оборский, В.М. Тонконогий, В.Д. Гогунский // Високі технології в машинобудуванні : зб. наук. праць. – 2014. – № 1 (24). – С. 130– 138. – doi.org\10.13140/RG.2.1.1405.6407.
3. Васильєва, В.Ю. Інтеграція у світову спільноту університетів через публікаційну активність в Internet – просторі / В.Ю. Васильєва, В.Д. Гогунський, Г.О. Оборский // Управління проектами: стан та перспективи : XII міжнар. наук.-практ. конф. – Миколаїв : НУК, 2016. – С. 12. doi: 10.13140/RG.2.1.4720.5360
4. Буй, Д. Б. Scopus та інші наукометричні бази: прості питання та нечіткі відповіді / Д. Б. Буй, А. О. Білощицький, В. Д. Гогунський // Вища школа. – 2014. – № 4. – С. 37 –40. – doi.org\10.13140/RG.2.1.1989.3205.
5. Бушуев, С. Д. Наукометричні бази: характеристика, можливості і завдання / С. Д. Бушуев, А. О. Білощицький, В. Д. Гогунський // Управління розвитком складних систем. – 2014. – № 18. – С. 145 –152. – doi.org\10.13140/RG.2.1.2196.9361
6. Загальні механізми формування системи цитування наукових статей / В.Д. Гогунський, ВА Яковенко, ТА Лященко, ТВ Отрадская // Вісник НТУ «ХПІ». Стратегічне управління. – 2016. – № 1 (1173). – С. 14 – 18. doi: http://dx.doi.org/10.13140/RG.2.1.1880.5203
7. Коляда, А. С. Автоматизация извлечения информации из наукометрических баз данных / А. С. Коляда, В. Д. Гогунский // Управління розвитком складних систем. – 2013. – № 16. – С. 96 – 99. doi.org\10.13140/RG.2.1.2668.7440
8. Негри, А. А. Концепция проекта агрегирующей аналитической информационной системы для работы с наукометрическими базами данных / А. А. Негри, Е. В. Колесникова, Ю.С. Барчанова // Інформаційні технології в освіті, науці та виробництві. – 2013. – № 4(5). – С. 52 – 56.
9. Білощицький А.О. Наукові засади застосування методів проектного менеджменту в векторних інформаційних технологіях управління підприємствами акредитації / А.О. Білощицький С.В. Білощицька // V між нар. наук.-практ. конф.

"Управління проектами: стан та перспективи." – Миколаїв : НУК. – 2009. – С. 129 – 130.

10. Копанєва, Є. О. Національні індекси наукового цитування / Є. О. Копанєва // Бібл. вісник. — 2012. — № 4. — С. 29 — 34.
11. Оборський, Г. О. Scopus: достовірність даних за запитами щодо числа публікацій університетів / Г. О. Оборський, В. Д. Гогунський, В. А. Волобоєв // Інформаційні технології в освіті, науці та виробництві : зб. – 2014. – № 2 (7). – С. 179 – 190. – doi.org\10.13140/RG.2.1.3384.7769
12. Яковенко, В.А. Scopus: поиск информации о публикациях ученых Одесского национального политехнического университета / В.А. Яковенко, А.А. Негри, Ю.С. Борчанова // Шляхи реалізації кредитно -модульної системи організації навчального процесу : наук.-метод. семінар. – 2014. – № 8. – С. 67 – 77.
13. Новиков, Д. А. Наукометрия и экспертиза в управлении наукой [Текст] / Д. А. Новиков, М. В. Губко // Упр. больш. сист. «Наукометрия и экспертиза в управлении наукой». — М. : ИПУ РАН, 2013. — Спец. вып. № 44. — С. 8—13.
14. Гогунський, В.Д. Особливості цитування наукових публікацій у Інтернет-просторі / В.Д. Гогунський, В.О. Яковенко, А.С. Коляда // Шляхи реалізації кредитно-модульної системи. – 2015. – № 10. – С. 28 – 33. – doi.org\10.13140/RG.2.1.5058.8885.
15. Коляда, А. С. Автоматизация извлечения информации из наукометрических баз данных / А. С. Коляда, В. Д. Гогунский // Управління розвитком складних систем. – 2013. – № 16. – С. 96 – 99. doi.org\10.13140/RG.2.1.2668.7440
16. Гогунский, В. Д. Наукометрические данные научного издания «Управление развитием сложных систем» / В. Д. Гогунский, А. С. Коляда, В. А. Яковенко // Управління розвитком складних систем. – 2014. – № 19. – С. 6 – 11. doi.org\10.13140/RG.2.1.3826.9847
17. Тернер, Дж. Родни Руководство по проектно-ориентированному управлению / Пер. с англ. под общ. ред. В.И. Воропаева. – М. : Изд. дом Гребенникова, 2007. – 552 с.
18. Бушуев, С.Д. Современные подходы к развитию методологий управления проектами / С.Д. Бушуев, Н.С. Бушуева // Управління проектами та розвиток виробництва: Зб. наук. праць.– Луганськ : Вид-во СНУ ім. В. Даля, 2005. – № 1(13). – С. 5 – 19.
19. Керівництво з управління інноваційними проектами та програмами. P2M . Том 1 , Версія 1.2: пров. з англ. / Під ред. проф. С.Д. Бушуєва. – К. : Наук. світ, 2009. – 173 с.

20. Оборський, Г. О. Стандартизація і сертифікація процесів управління якістю освіти у вищому навчальному закладі [Текст] / Г. О. Оборський, В. Д. Гогунський, О. С. Савельєва // Тр. Одес. политехн. ун-та. – Вып. 1(35). – 2011. – С. 251 – 255.
21. Ткачук, С.В. Профілювання цінності проектів освітньої діяльності для навчальних закладів / СВ Ткачук, ВД Гогунський // Шляхи реалізації кредитномодульної системи організації навчального процесу ... – 2011. – № 4 (5). – С. 58 – 63.
22. Ткачук, С. В. Багатовекторний розвиток навчальних закладів на основі концепції створюваної цінності / С. В. Ткачук, В. Д. Гогунський // Інформ. технології в освіті, науці та виробництві. – 2013. – № 1 (2). – С. 256 – 260. doi: doi.org\10.13140/RG.2.1.2401.7364
23. Романенко, Н.В. Определение ценности проектов в здравоохранении / Н.В. Романенко, С.В. Руденко, А.В. Шахов // Вісник Одеського нац. морськ. ун-ту.: зб. наук. праць. – Одеса, ОНМУ, 2010. – Випуск 31. – С. 162 – 171.
24. Белощицкий, А. А. Управление проблемами в методологии проектно-векторного управления образовательными средами [Текст] / А. А. Белощицкий // Управління розвитком складних систем. – 2012. – № 9. – С. 104 – 107.
25. Гогунский, В.Д. Розробка моделі життєвого циклу наукових публікацій / В.Д. Гогунський, Т.О. Лященко, В.Ю. Васильєва // Управління розвитком складних систем. – 2015. – № 24. – С. 75 – 83. – doi.org\ 10.13140/RG.2.1.4442.8564
26. Gogunsky, V.D. Scientometric data scientific publication «Management of development of complex systems» / V.D. Gogunsky, A.S. Kolyada, V.O. Iakovenko // Management of development of complex systems. – 2014. – № 19. – PP. 6 – 11
27. Бушуев, С.Д. Современные подходы к развитию методологий управления проектами / С.Д. Бушуев, Н.С. Бушуева // Управління проектами та розвиток виробництва: Зб. наук. праць.– Луганськ : Вид – во СНУ ім. В. Даля, 2005. – № 1(13). – С. 5 – 19.
28. Яковенко А.Е Стратегия принятия решений в условиях адаптивного обучения / Яковенко А.Е. Нарожный А.В. Гогунский В.Д. // Восточно-европейский журнал передових технологий. – 2/2(14). – 2005. – С.105 – 110
29. Бушуев С.Д. Управление проектами развития от видения к реальности // Міжнар. конф. «Управління проектами у розвитку суспільства». – К. : КНУБА, 2005. – С. 15 – 18.
30. Оборський, Г. О. Нові тенденції і завдання щодо підготовки науковців вищої кваліфікації [Текст]/ Г. О. Оборський, В. Д. Гогунський Інформ. технології в освіті, науці та виробництві : зб. наук. праць. – Вип. 2. – Одеса : АО Бахва, 2013 – С. 15 – 22.

31. Гогунський, В. SCOPUS: знайдемо свої публикации / В. Гогунський, Д. Буй // Вища школа. – 2014. – №8 (121–122). – С. 113 – 115.
32. Управление проектами повышения публикационной активности в информационных интернет-ресурсах / В.А. Яковенко; В.Ю. Васильева; А.С. Коляда; В.Д. Гогунский // Інформаційні технології та взаємрдії. III міжнар. наук.-практ. конф. – Київ : КНУ ім. Тараса Шевченка, 2016. – С. 135 – 136.
33. Оборський, Г.О. Наукометрические исследования публикационной активности, как составляющая инновационного развития университета / Г.О. Оборський, В.М. Тонконогий, В.Д. Гогунський // Високі технології: тенденції розвитку. Матер. XXIII міжнар. наук.-техн. семінару, 7–12 вересня 2015 р., м. Одеса.– С. 126 – 127.
34. Gogunsky V.D. The development of the system concept of scientometric databases / V.D. Gogunsky, V.O. Iakovenko, A.S. Kolyada // Management of Development of Complex Systems. – 2014. – № 20. – pp. 143 – 147.
35. Гогунський, В. Д. Наукометричні бази: характеристика, можливості і завдання / В. Д. Гогунський, Г.О. Оборський, А. С. Коляда // Наук.-метод. семінар: „Шляхи реалізації кредитно-модульної системи". – Вип. 8. — Одеса : Наука і техніка, 2014. — С. 3 — 12.
36. Гогунський, В.Д . Особливості цитування наукових публікацій у інтернет-просторі / В.Д. Гогунський, В.О. Яковенко, А.С. Коляда // Наук.-метод. семінар: „Шляхи реалізації кредитно-модульної системи ". – Вип. 10. — Одеса : Наука і техніка, 2015. — С. 28 – 33.
37. Коляда, А. С. Вилучення інформації із слабоструктурованих веб сторінок / А. С. Коляда, В. Д. Гогунський // Східно-Європ. журнал передових технологій. – № 1/9 (67). – Харків : Технолог. центр, 2014 – С. 51 – 54.
38. Harzing, Anne–Wil. The Publish or Perish Book. – Tarma Software Research Pty Ltd, Мельбурн, Австралия. – 2010. – 266 с.
39. Hirsch, J. E. An index to quantify an individual's scientific research output [Текст] // arXiv: physics/0508025. – v5. – 29 Sep. 2005. – 5 p.
40. Гогунський, В. Створюємо свій акаунт "GOOGLE Академія" [Текст] / В.Д. Гогунський, О.Є. Колесніков // Вища школа. – 2014. – № 9. – С. 55 –58. – doi.org\10.13140/RG.2.1.3253.9609.
41. Гогунський, В. SCOPUS: Пошук статей за прізвищем автора / Віктор Гогунський, Андрій Білощицький // Вища школа. – 2015. – № 3–4. – С. 115 – 117. *doi.org\10.13140/RG.2.1.1740.1680*

42. Коляда, А. С. Латентно семантичний підхід для аналізу інформації із наукометрических баз даних [Текст] / А. С. Коляда // Управління розвитком складних систем. – 2014. – Вып. 17. – С. 90 – 94.
43. Коляда, А. С. Достовірність ідентифікації авторства научних публікацій на основі латентно семантичного аналізу [Текст] / А. С. Коляда, В. Д. Гогунський // Східно-Європ. журнал передових технологій. – № 3/2 (69). – Харків : Технолог. центр, 2014 – С. 36 – 40.
44. Коляда, А. С. Латентно семантический аналіз информації із наукометрических баз. / А. С. Коляда // Наук.-метод. семінар: „Шляхи реалізації кредитно-модульної системи ”. – Вип. 9. — Одеса : Наука і техніка, 2014. — С. 30 – 36.
45. О'Шонесси, Дж. (O'Shaughnessy John). Конкурентный маркетинг. Стратегический подход. – С-Пб.: Питер, 2002. – 864 с.
46. Мазур, И.И., Управление проектами [текст] / И.И. Мазур, В.Д. Шапиро, Н.Г. Ольдерогге; под общей ред. проф. Мазура И.И. – М.: Экономика, 2001. – 574 с.
47. Вейл, П. Лидерство, основанное на видении. КурсMBA по менеджменту. – М., 2004. – 338с.
48. Бушуев, С.Д. National Competence Baseline, NCB UA Version 3.1 [Text] / С.Д. Бушуев, Н.С. Бушуева. – К. : ІРІДІУМ, 2010. – 208 с.
49. Про затвердження Порядку присудження наукових ступенів і присвоєння вченого звання старшого наукового співробітника. Постанова КМУ № 567 від 24.07.13 р. – http://osvita.ua/legislation/Vishya_osvita/36856/.
50. Про теми дисертаційних робіт. Лист МОНмолодьспорту України від 14.02.2013 № 1/9-116 – http://mon.gov.ua/ua/activity/certified-staff-evaluation/564 - *23.02.2013*.
51. Про затвердження орієнтовних критеріїв оцінювання діяльності вищих навчальних закладів. – Наказ МОН України від 20.06.1013 р. № 809.
52. Про внесення змін до наказу Міністерства освіти і науки, молоді та спорту України від 17 жовтня 2012 року № 1112 «Про опублікування результатів дисертацій на здобуття наукових ступенів доктора і кандидата наук». – Наказ МОНмолодьспорту України від 3.12.2013 № 1380. – *http://mon.gov.ua/ua/activity/certified-staff-evaluation/564/*.
53. Оборський, Г.О. Нові тенденції і завдання щодо підготовки науковців вищої кваліфікації [Текст] / Г.О. Оборський, В.Д. Гогунський // Інформаційні технології в освіті, науці та виробництві. – Вип. 2 (5). – О. : АО Бахва, 2013. – С. 15 – 22.
54. Костирко, Т. Н. Університети України: приєднання до руху відкритого доступу // Вісник ОНУ. – Том 16. – Випуск 1/2 (5/6). – 2011. – С. 283 – 289.

55. Гогунський, В .Д. Разработка концепции системы наукометрической базы даннях / В.Д. Гогунський, В.А. Яковенко А.С. Коляда // Управління розвитком складних систем. – 2014. – № 20. – С. 143 – 147.
56. Шанхайский рейтинг лучших университетов мира ARWU (Электронный ресурс) – Режим доступа: http://www.shanghairanking.com/ru/
57. Рейтинг лучших университетов мира по версии QS [Электронный ресурс] — Режим доступа: http://gtmarket.ru/ratings/qs-world-university-rankings/info.
58. Рейтинг лучших университетов мира Times Higher Education (THE) (Электронный ресурс) – Режим доступа: https://www.timeshighereducation.co.uk/world-university-rankings/
59. Рейтинг UI GreenMetric Ranking of World Universities. (Електронний ресурс) – Режим доступу: http://greenmetric.ui.ac.id/ranking/
60. Белл Д. Грядущее постиндустриальное общество: опыт социального прогнозирования / Д. Белл; пер. с англ. / Под ред. В.Л. Иноземцева. – М.: Academia, 1993. – С. 28 – 118.
61. Вайсман, В. О. Сучасна концепція проектно-орієнтованого командного управління підприємством / В. О. Вайсман, К. В. Колеснікова, В. В.Натальчишин // Сучасні технології в машинобудуванні: зб. наук. праць. – 2013. – Вип. 8. – НТУ «ХПІ». – С. 246 – 253..
62. Колеснікова, К. В. Моделювання стратегічного управління міжнародною діяльністю університету [Текст] / К.В. Колеснікова, С.М. Гловацька, С.В. Руденко // Проблеми техніки. – № 1. – 2013. – С. 95 – 101.
63. Рач, В. А. Контекстно-личностное оценивание компетентности проектных менеджеров с использованием теории нечетких множеств [Текст] / В.А. Рач, О.В. Бирюков // Управління проектами та розвиток виробництва: зб. наук. пр. – Луганськ : СНУ ім. В. Даля. 2009. – № 1 (29). – С. 151 – 169.
64. Бушуев, С. Д. Напрями дисертаційних наукових досліджень зі спеціальності «Управління проектами та програмами» [Текст] / С.Д. Бушуев, В. Д. Гогунський, К.В. Кошкін // Управління розвитком складних систем. – 2012. – № 12. – С. 5 – 7.
65. Арчибальд, Р. Управление высокотехнологичными программами и проектами [Текст] / Рассел Д. Арчибальд; пер. с англ. Мамонтова Е.В.; Под ред. Баженова А.Д., Арефьева А.О. – 3-е изд. – М. : Компания АйТи; ДМК Пресс, 2004. – 472 с.
66. Долгосрочный прогноз социально-экономического развития Украины: монография / И.В. Кононенко, В.Л. Лисицкий, А.С. Пономарев та ін. / Под общ. ред. И.В. Кононенко. – Харьков : ИМиС, 1999. – 176 с.

67. Bushuyev, Sergey D. Entropy measurement as a project control tool international / Sergey D. Bushuyev, Sergey V. Sochnev // Journal of Project Management. – Elsevier, 1999. – 17 (6). – P. 343 – 350.
68. Кошкин, К.В. Оценка эффективности портфеля проектов судостроительного предприятия / К.В. Кошкин, А.М. Возный // Зб. наук. праць Нац. ун-ту кораблебудування. – Миколаїв : НУК, 2006. – С. 10 – 13.
69. Кононенко, И.В. Математическая модель и метод оптимизации содержания проекта с точки зрения времени и стоимости его выполнения [Текст] / И.В. Кононенко, В.А. Мироненко // Вост.-Европ. журнал передовых технологий. – Харьков: Технол. центр, 2010. – № 1/2 (43). – С.12 – 17.
70. Креативные технологии управления проектами и программами / С.Д.Бушуев, Н.С.Бушуева, И.А. Бабаев та ін. – К.: Саммит – Книга, 2010. – 768 с.
71. Гогунський, В.Д. Управління процесом формування наукометричних показників наукових публікацій / В.Д. Гогунський, В.Ю. Васильева, В.О. Яковенко // Інформ. технології в освіті, науці та виробництві : зб. наук. праць. – Вип. 4 (11). – Одеса : АО Бахва, 2015. – С. 6 – 18.
72. Peter Buneman, Semistructured data, Proceedings of the sixteenth ACM SIGACT–SIGMOD–SIGART symposium on Principles of database systems, p.117 – 121, May 11 – 15, 1997, Tucson, Arizona, United States.
73. Arens, Yigal. Retrieving and integrating data from multiple information sources / Yigal Arens, Chin Y. Chee, Chun-Nan Hsu, Craig A. Knoblock // International Journal of Intelligent and Cooperative Information Systems. Issue 02 – 1993.
74. Yung-Jen Hsu, Jane. Template-based information mining from HTML documents / Jane Yung-Jen Hsu, Wen-tau Yih // Proceedings of the fourteenth national conference on artificial intelligence and ninth conference on Innovative applications of artificial intelligence. – 1997, p. 256 – 262.
75. Smith, Dan. Information extraction for semi-structured documents / Dan Smith, Mauricio Lopez // In Proceedings of the Workshop on Management of Semistructured Data. – 1997.
76. Li, Zhao. Web data extraction based on structural similarity / Zhao Li, Wee Keong Ng, Aixin Sun // Journal Knowledge and Information Systems archive Volume 8 Issue 4, November 2005, p. 438 – 461.
77. Коляда А. С. Разработка проекта информационно-аналитической системы извлечения и обработки информации из наукометрических баз данных / Коляда А. С., Негри А. А., Колесникова Е. В. // Управління проектами: стан та перспективи. Матеріали IX Міжнар. наук.-практ. конф. — Миколаїв : НУК, 2013. — 348 с.

78. Білощицький, А.О. Наукометричні бази та індикатори цитування наукових публікацій / А.О. Білощицький, В.Д. Гогунський // Інформаційні технології в освіті, науці та виробництві. – Вип. 4 (5). – О. : АО Бахва, 2013. – С. 198 – 203.
79. Robert Baumgartner. The Personal Publication Reader: Illustrating Web Data Extraction, Personalization and Reasoning for the Semantic Web / Robert Baumgartner, Nicola Henze, Marcus Herzog // Lecture Notes in Computer Science Volume 3532, 2005, pp 515–530
80. Формализация проблемы извлечения знаний из естественно языковых текстов. [Текст] / [А. Палагин, С. Кривый, Н. Петренко, Д. Бибиков]. — Sofia. : Information technologies & knowledge, 2012. — 100 с.
81. Scopus (Elsevier): Elsevier receives millionth response to Editor, Author and Reviewer . – *http://www.scopus.com/search/form/authorFreeLookup.url*
82. РИНЦ – Научная электронная библиотека. – http://www.elibrary.ru
83. Bielefeld Academic Search Engine. – http://www.base-search.net
84. Index Copernicus. Indeksacja czasopisma. – http://www.journals.indexcopernicus.com/search_article.php
85. Springer Science + Business Media. – http://www.springer.com
86. Scrapy – a fast high–level screen scraping and web crawling framework. – http://scrapy.org
87. MongoDB – an open–source document database. – *http://ru.wikipedia.org/wiki/MongoDB*
88. Белощицкий А.А. Модель расширяющейся вселенной проектов в управлении образовательными средами / А.А. Белощицкий // Восточно-Европейский журнал передовых технологий. – 2012. – № 1/11 (55). – С. 41–43.
89. Білощицький А.О. Інформаційна технологія управління проектами на базі ERPP (enterprise resources planning in project) та APE (administrated projects of the enterprise) систем / Ю.М. Тесля, А.О. Білощицький, Н.Ю. Тесля // Зб. наук. пр.: Управління розвитком складних систем. – Вип. 1. – К. : КНУБА, 2010. –С.16–20.
90. Чугреев, В. Л. Модель структурного представления текстовой информации и метод ее тематического анализа на основе частотно-контекстной классификации / Санкт-Петербургский гос. электротехнический ун-т "ЛЭТИ" им. В.И. Ульянова. 2003. – С. 25 29.
91. Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, Richard Harshman (1990). Indexing by Latent Semantic Analysis. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE. 41(6):391–407.

92. Řehůřek, R. (2011). Subspace tracking for latent semantic analysis. Advances in Information Retrieval. 289–300
93. Высоцкий, В.Ю. Разработка обучающих программ в виртуальной компьютерной среде // В.Ю. Высоцкий, В.Д. Гогунский // Тр. Одес. политехн. ун-та. – Вып. 2 (36). – 2011. – С. 184 – 189.
94. Яковенко, В.Д Комп'ютерна реалізація системи автоматизованого управління навчальним процесом // В. Д. Яковенко, В. Д. Гогунський, Г. Ф. Сафонова // Моделир. в прикладных научных исследованиях. Матер. XVI семинара. — Одеса : ОНПУ, 2008. – С. 27 – 30.
95. Тертышная, Т. И. Автоматизированная система контроля знаний / Т. И. Тертышная, Е. В. Колесникова, В. Д. Гогунский // Тр. Одес. политехн. ун-та. — Вып. 1(13).—2001. — С. 125 — 128.
96. Яковенко, А.Е Стратегия принятия решений в условиях адаптивного обучения / А. Е. Яковенко, А. В. Нарожный, В. Д. Гогунский // Восточно-европейский журнал передових технологий. – № 2/2 (14). – 2005. – С.105 – 110.
97. Porter M.F. (1980). An algorithm for suffix stripping. Program, 14 no. 3, pp 130–137.
98. Гогунский, В.Д. Обоснование закона о конкурентных свойствах проектов / В. Д. Гогунский, С.В. Руденко, П.А. Тесленко // Управління розвитком складних систем. – № 8. – 2012. – С. 14 – 16.
99. Рач, В. А. Побудова термінологічної системи організації наукового знання / В. Рач, О. Россошанська, О. Медведєва // Науковий світ. – 2011. № 4. – С. 13 – 16.
100. Гогунський, В. Д. Марковські моделі комунікаційних процесів в міжнародних проектах / О. В. Власенко, В. В. Лебідь, В. Д. Гогунський // Управління розвитком складних систем. № 12. – 2012.– С. 35 – 39.
101. Плетнев, А.Н. Организация вычислительной сети студгородка «Политехник» с использованием оптического волокна / А.Н. Плетнев, А.Н. Миколюк, В.Д. Гогунский // Труды Одес. политехн. ун-та. – 2007. – № 2(28). – С. 138 – 140.
102. Колесникова, Е. В. Моделирование слабо структурированных систем проектного управления / Е. В. Колеснікова // Тр. Одес. политехн. ун-та. - 2013. – № 3 (42). - С. 127 – 131. - doi.org\10.15276/opu.3.42.2013.25
103. Колесникова, Е. В. Развитие теории проектного управления: закон Ю.Л. Воробьева о влиянии риска на успешность портфеля проектов / Е. В. Колесникова // Управління розвитком складних систем. – 2014. – № 18. – С.62 – 67.
104. Вайсман, В.О. Моделі, методи та механізми створення і функціонування проектно-керованої організації : Монографія / В.О. Вайсман. – К. : Наук. світ, 2009. – 146 с.

105. Фарионова, Т. .А. Когнитивное моделирование в проектировании композиционных материалов и покрытий / Т .А .Фарионова, Ю .А. Казимиенко // Вост.-Европейский журнал передовых технологий. – 2011. – 1/6 (49). – С. 36 – 38.
106. Де Боно, Эдвард. Шесть шляп мышления. – Минск: Попурри, 2006. – 208 с.
107. Ма Фен. Марковская модель процесса формирования и управления имиджем учебного заведения / Ма Фен, С.Н. Гловацкая, Е.В. Колесникова // Проблеми техніки. – 2013. – № 3. – С. 142 – 151.
108. Руденко, С.В. Анализ результатов реализации технико-экономической природоохранной региональной программы / С.В. Руденко, Е.В. Колесникова, Т.М. Олех // Проблеми техніки. – № 2. –2013. – С. 161 – 169.
109. Толмен, Э. Когнитивные карты у крыс и у человека [Электронный ресурс]. – Хрестоматия по истории психологии. Под ред. Гальперина П. Я. – М. : Изд-во МГУ, 1980. – С. 63 – 69. – http://www.psychology.ru/library/00060.shtml
110. Кошкин, К.В. Когнитивные модели управления жилищно-коммунальным хозяйством как активной системой / К.В. Кошкин, С.А. Макеев, Г.В. Фоменко // Управління розвитком складних систем – 2011. –№ 5. – С. 17 – 19.
111. Колесникова, Е.В. Трансформация когнитивных карт в модели марковских процессов для проектов создания программного обеспечения / Е.В. Колесникова, А.А. Негри // Управління розвитком складних систем. – №15. – 2013. – С. 30 – 35.
112. Колесникова, Е.В. Когнитивные модели слабо структурированных проектов создания программных продуктов // Моделир. в прикл. научных исследованиях : Матер. XX семинара. – Одесса : ОНПУ, 2012. – С. 48 – 50.
113. Власенко, О.В. Марковські моделі комунікаційних процесів в міжнародних проектах / О.В. Власенко, В.В. Лебідь, В.Д. Гогунський // Управління розвитком складних систем. – 2012. – № 12. – С. 35 – 39.
114. Gogunsky, V.D. Markov model of risk in projects of safety / V.D. Gogunsky, Yu.S. Chernega, E.S. Rudenko // Тр. Одес. политехн. ун-та. – Вып. 2 (41). – 2013. – С. 271 – 276.
115. Власенко Е.В. Модель «Діамант» оцінки внутрішніх комунікацій в Європейських проектах [Текст] / Е.В. Власенко, Д.В. Лукьянов, В.Д. Гогунский // Вост.-Европ. журнал передовых технол. – № 1/10 (61). – Харьков: Технолог. центр, 2013. – С. 86 – 88.
116. Колесникова, Е.В. Управление знаниями в IT-проектах / Е.В. Колесникова, А.А. Негри // Вост.-Европ. Журнал передовых технологий. – 2013. – № 1/10 (61). – С. 213 – 215.

117. Бондарь, В.И. Проявление закона Кошкина КВ в безнадежных проектах: признаки, свойства, результаты / В.И. Бондарь, В.Д. Гогунский // Управління проектами: стан та перспективи: конф. – 2009. - С. 111-112
118. Коджа, Т.И. Определение необходимых и достаточных условий объективности оценки результатов тестирования / Т.И. Коджа, В.Д Гогунский // Тр. Одес. политехн. ун-та. - 2002.-Спецвыпуск. – С. 87-88
119. Тесленко, П.А. Эволюционная парадигма проектного управления / П.А. Тесленко, В.Д. Гогунский // Управління проектами: Стан та перспективи. VI МНПК 6. – 2010. – С. 114 - 117
120. Oganov, A.V. Using the theory of constraints in implementing enterprise project management office / A.V. Oganov, V.D. Gogunsky // GESJ: Computer Sciences and Telecommunications. – 2013. - № 4 (40). – P. 59-65
121. Запорожець, О.І. Завдання наукових досліджень з охорони праці / О.І. Запорожець, В.Д. Гогунський // Інформ. технології в освіті, науці та виробництві. – 2013. - № 4 (5). - С. 19 – 23.
122. Риск сокращения продолжительности жизни: рабочая зона / ЕЕ Басиль, СА Изотов, ВД Гогунский // Тр. Одес. политехн. ун-та. - 1997. - № 2 (2). - С. 133-135.
123. Коляда, А. С. Применение латентного размещения Дирихле для анализа публикаций из наукометрических баз данных / А.С. Коляда, В.А. Яковенко, В.Д. Гогунский // Тр. Одес. политехн. ун-та – 2014. - № 1 (43). - С. 186-191. doi: http://dx.doi.org\10.15276/opu.1.43.2014.32
124. Коляда, А.С. Извлечение информации из слабоструктурированных веб-страниц / А. С Коляда, В.Д. Гогунский // Eastern-European Journal of Enterprise Technologies. 2014, № 1/9 (67), С. 51-54 dx.doi.org/10.15587/1729-4061.2014.19496
125. Руденко, С.В. Оценка экологической безопасности в проектах / СВ Руденко, ВД Гогунский // Монография. – 2006. – 144 с.

Дисертації, що захищені виконавцями роботи за тематикою досліджень:

126. Становская, И.И. Балансирование и гармонизация решений в управлении программами, состоящими из серийных проектов: дисс. … канд. техн. наук: 05.13.22 / Становская Ираида Ивановна [Науч. рук., к.т.н., доц. Колесникова Е.В.]. – Одесса : ОНПУ, 2013. – 203 с.
127. Власенко, О.В. Управління комунікаціями у міжнародних проектах в рамках європейських програм: дис. … канд. техн. наук: 05.13.22 / Власенко Олена Вікторівна [Наук. керівн., д.т.н. Гогунський В.Д.]. – Одеса : ОНПУ, 2014. – 187 с.

128. Лукьянов, Д.В. Модели и методы управления знаниями в проектах на основе компетентностного похода: дисс. … канд. техн. наук: 05.13.22 / Лукьянов Дмитрий Владимирович [Науч. рук., к.т.н., доц. Колесникова Е.В.]. – Одесса : ОНПУ, 2014. – 202 с.
129. Олех, Т.М. Разработка моделей целеполагания и методов принятия решений в проектах на основании многомерных оценок: дисс. … канд. техн. наук: 05.13.22 / Олех Татьяна Мефодиевна [Науч. рук., д.т.н., проф. Гогунский В.Д.]. – Одесса : ОНПУ, 2015. – 150 с.
130. Яковенко, Є.О. Моделі та методи експертного оцінювання рівня корпоративних знань для прийняття проектних рішень: дис. … канд. техн. наук: 05.13.22 / Яковенко Євген Олександрович [Наук. керівн., д.т.н., проф. Гогунський В.Д.]. – Одеса : ОНПУ, 2015. – 137 с.
131. Коляда, А.С. Моделі і методи пошуку інформації у наукометричних базах даних: дис. … канд. техн. наук: 05.13.06 / Коляда Андрій Сергійович [Наук. керівн., д.т.н., проф. Гогунський В.Д.]. – Одеса : ОНПУ, 2015. – 113 с.
132. Москалюк, А.Ю. Моделі і методи управління ініціацією проектів охорони праці: дис. … канд. техн. наук: 05.13.22 / Москалюк Андрій Юрійович [Наук. керівн., д.т.н., проф. Гогунський В.Д.]. – Одеса : ОНПУ, 2016. – 142 с.
133. Колеснікова, К.В. Методологія структурного та параметричного аналізу систем проектного управління: дис. … д-ра техн. наук: 05.13.22 / Колеснікова Катерина Вікторівна [Наук. конс., д.т.н., проф. Руденко С.В.] – Миколаїв : НУК, 2015. – 313с.

Публікації виконавців за тематикою НДР: статті у журналах, що входять до наукометричних баз даних SCOPUS або WoS.

134. Development of the model of interaction among the project, team of project and project environment in project system / O. Kolesnikov, V. Gogunskii, K. Kolesnikova, D. Lukianov, T. Olekh // Eastern-European Journal of Enterprise Technologies. – 2016. – № 5/9 (83). – С. 20 – 26 DOI: http://dx.doi.org\10.15587/1729-4061.2016.80769
135. "Lifelong learning" is a new paradigm of personnel training in enterprises / V. Gogunskii, A. Kolesnikov, K. Kolesnikova, D. Lukianov // Eastern-European Journal of Enterprise Technologies. – 2016. – № 4/2 (82). – P. 4–10. DOI: 10.15587/1729-4061.2016.74905
136. Development of parametric model of prediction and evaluation of the quality level of educational institutions / T. Otradskaya, V. Gogunskii, S. Antoschuk, O. Kolesnikov // Eastern-European Journal of Enterprise Technologies. – 2016. – 5/3 (83). – P. 12-21. DOI: http://dx.doi.org/10.15587/1729-4061.2016.80790

137. Development process models for evaluation of performance of the educational establishments / T. Otradskaya, V. Gogunskii // Eastern-European Journal of Enterprise Technologies. – 2016. – № 3 (3/81). – P. 12 – 22. DOI: 10.15587/1729–4061.2016.66562
138. Пройдак, Ю.С. Підвищення якості вищої освіти шляхом формування систми критеріїв розвитку ВНЗ / Ю.С. Пройдак, В.В. Малий, В.М. Молоканова, К.В. Колеснікова // International scientific Journal Acta Universitatis Pontica Euxinus – Special number: XII International conference «Strategy of quality in industry and education», 2016, Varna, Bulgaria. – С. 432 – 438 WoS
139. Sherstyuk, O. The research on role differentiation as a method of forming the project team / O. Sherstyuk, T. Olekh, K. Kolesnikova // Eastern–European Journal of Enterprise Technologies. – 2016. – № 2/3 (80). – P. 63–68. DOI: http://dx.doi.org/10.15587/1729–4061.2016.65681]
140. Kolesnikova, K. Experimental and analytical description of the electric arc furnace processes in creation of computer simulator // Metallurgical and Mining Industry. – 2015. – № 12. – P. 55–59. DOI: doi.org\10.13140/RG.2.1.4602.8881
141. Dynamic models in the method of project management / A. Stanovsky, K. Kolesnikova, E. Lebedeva, I. Kheblov // Eastern–European Journal of Enterprise Technologies. – 2015. – № 6/3 (78). – P. 46–52 DOI: https://doi.org/10.15587/1729–4061.2015.55665