

Голові спеціалізованої вченої ради К58.052.06
Тернопільського національного технічного
університету імені Івана Пулюя

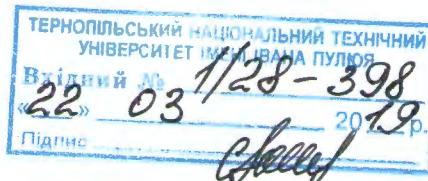
46001, м. Тернопіль, вул. Руська, 56.

ВІДГУК

офіційного опонента – завідувача спеціальної кафедри № 5 Інституту спеціального зв’язку та захисту інформації Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», доктора технічних наук, доцента Субача Ігоря Юрійовича на дисертацію Андрушенко Валентини Борисівни «Інформаційні технології наукометричного аналізу на основі моніторингу ресурсів мережі Інтернет», подану на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології

Актуальність теми дисертації. У теперішній час інформаційні технології все більше стають основним інструментом для обробки та аналізу наукометричної інформації. Саме розвиток і використання інформаційних технологій дозволяє сформувати всебічний погляд на аналіз публікаційної активності і надає можливість подальшого використання отриманих даних для прийняття важливих управлінських рішень.

Дані технології орієнтовані на обробку інформації, що переважно міститься в мережі Інтернет. Кожен інформаційний ресурс, що містить інформацію про науковий журнал, публікації, авторів, тощо, відрізняється один від одного не тільки дизайном (організація меню, ілюстративний матеріал, тощо) та доступом до інформації (відкритий чи передплачений), а й наборами даних, що містить певний ресурс і способами пошуку та представлення інформації, можливостями збереження отриманих в результаті пошуку даних та ін.



Нажаль, на сьогодні більшість наукових робіт щодо формування та оцінки роботи інформаційних систем, в основному, зосереджена на питаннях аналізу освітніх ресурсів з науковою інформацією та питаннях її захисту.

Таким чином, у теперішній час **актуальним науковим завданням** залишається розробка інформаційних технологій для проведення наукометричного аналізу на основі моніторингу ресурсів мережі Інтернет, формування масивів інформації, розробки відповідних моделей і алгоритмів реалізації додаткових можливостей для існуючих наукометричних ресурсів та ресурсів наукової інформації відкритого доступу.

Наукова новизна. Автором дисертації отримано наукові результати, що в сукупності вирішують актуальне наукове завдання. Найбільш суттєвими з них є наступні.

1. Вперше розроблено моделі та алгоритми побудови предметних областей на базі ресурсу Google Scholar Citations, наукова новизна полягає у використанні тематичних дескрипторів для обмеження розмірів мереж.

2. Вперше запропоновано критерії аналізу наукометричних ресурсів за показниками повноти та доступності інформації для користувача, шляхом їхнього представлення у вигляді логічної матриці.

3. Вперше запропонована і реалізована на основі енциклопедичного ресурсу Вікіпедія нова інформаційна технологія: а) обчислення індексу популярності автора (науковця), що дозволяє здійснювати моніторинг популярності науковця в рамках джерел науково-популярної інформації; б) побудови онтології понять, що дозволяє автоматизувати процес пошуку і формування онтології для заданого поняття для подальшого визначення зв'язків між окремими поняттями.

Практичне значення одержаних автором наукових результатів полягає у можливості реалізації та використанні на практиці додаткового до традиційних наукометричних показників при аналізі результативності вчених, колективів науковців, наукових установ та результативності реалізації конкурсних проектів.

Крім того, запропоновані моделі дозволять виокремити міждисциплінарну складову, що розшириТЬ межі сприйняття визначеного поняття, з точки зору приналежності до певної предметної області, для залучення фахівців різних галузей знань для реалізації глобальних проектів. Важливою є можливість використання отриманих результатів з точки зору охоплення даних не тільки ресурсів звичних і доступних лише науковому загалу, а й тих, що використовуються широким колом користувачів глобальних мереж.

Результати дослідження було використано при формуванні стратегії підготовки грантового дослідження Підприємством Української академії наук «Інститут системних досліджень та інформаційних технологій» та при формуванні матеріалів щодо розвитку навичок студентів та викладачів кафедри прикладної математики НТУУ «КПІ імені Ігоря Сікорського» основним елементам підготовки грантових запитів у межах Рамкової програми ЄС з досліджень та інновацій «Горизонт 2020» (акт від 14.01.2019р.).

Повнота викладу основних результатів дисертації в публікаціях.

Основні результати дисертаційної роботи з необхідною повнотою відображені у 7 наукових публікаціях, із яких 5 наукових статей у фахових виданнях, у тому числі, 2 статті у виданнях, які індексуються у міжнародних наукометрических базах, 14 тезах доповідей на наукових конференціях та 2 публікаціях в архіві препринтів.

Оцінка змісту дисертації, відповідність встановленим вимогам щодо оформлення. За своїм змістом дисертація Андрушченко В.Б. відповідає діючим вимогам щодо дисертацій на здобуття наукового ступеня кандидата наук і являє собою наукову працю, яка містить сукупність наукових результатів, виставлених автором для публічного захисту, має внутрішню єдність і свідчить про особистий внесок автора у науку.

Оформлення дисертації та автореферату відповідає вимогам Державних стандартів України. Текст дисертації та автореферату написані ясно та зрозуміло.

У *вступі* обґрунтовано актуальність теми дисертації, сформульовано мету, об'єкт, предмет, завдання дослідження, наукову новизну одержаних результатів, практичне значення результатів, зв'язок роботи з науковими програмами, планами, темами досліджень. Визначено особистий внесок здобувача, відомості про апробацію результатів роботи, публікації.

У *першому розділі* дисертації проведено огляд літературних джерел, пов'язаних із темою дослідження. Проведено аналіз основних аспектів, що відображені у науковій літературі щодо наукової інформації та ресурсів наукової та наукометричної інформації.

Описано засади, що ілюструють перелік питань, які цікавлять дослідників щодо можливостей обробки інформаційних масивів, які представлені ресурсами відкритого доступу та передплачених ресурсів.

Представлено модель користувача інформаційної системи наукової інформації. На основі моделі запропоновано низку критеріїв щодо оцінки ресурсів наукової інформації.

Другий розділ дисертації присвячений опису моделі інформаційного ресурсу наукової інформації. Проведений аналіз масиву ресурсів наукової інформації, здійснено огляд сучасних реферативних, наукометрических

ресурсів, репозитаріїв, архівів препринтів. За рахунок проведеного аналізу запропоновано інструментарій для розробки систем наукової інформації та наукометричних ресурсів. При цьому, основну увагу приділено ресурсам, на базі яких було реалізовано розробку додаткових інструментів для розширення можливостей і формування даних для їх подальшого використання та інтерпретації.

Третій розділ дисертації присвячений розробці методів та алгоритмів, для досягнення мети дослідження та описано підходи до реалізації часткових задач, що сформульовані у дисертаційній роботі.

Серед методів, що були застосовані при формуванні моделей та реалізації поставлених в роботі задач: методи комп’ютерної лінгвістики, методи статистичного аналізу даних, кількісні методи наукометричного аналізу, теорія графів, теорія складних мереж та інші.

Згідно із завданнями було проведено графометричний аналіз тексту, який містився на веб-сторінках ресурсів, який полягає у виокремленні слів та словосполучень, які відповідають: для системи Google Scholar - лексичним одиницям, що зображують теги (наукові інтереси, наукові напрямки користувача-науковця, які науковець зазначає самостійно у системі при реєстрації та оновленні профілю); для он-лайн енциклопедії Вікіпедія було опрацьовано корпуси текстів і виокремлено в текстових масивах власні імена та гіперпосилання для подальшої обробки отриманої інформації та її візуалізації; для ресурсу препринтів arXiv було проведено складання словників наукових напрямків для кожної з предметних областей за якими систематизовано публікації.

Було реалізовано метод та алгоритм побудови мережі співавторів на базі ресурсу Google Scholar Citations, який може бути допоміжним засобом при побудові складних наукових колаборацій, визначені експертів для вирішення різних задач. В основу методу було покладено теорію складних мереж, а для побудованої мережі було обчислено найкоротший шлях, що представляє собою мінімальну відстань між вузлами і обраховується як середня за всімаарами вузлів мінімальна відстань між ними.

Також було запропоновано метод та реалізовано алгоритм побудови онтології поняття на базі енциклопедичного ресурсу відкритого доступу Вікіпедія, який передбачає врахування гіперпосилань на статті з тих чи інших понять при побудові мережі.

Для побудови мережі предметних областей на базі ресурсу препринтів arXiv, було запропоновано метод та реалізовано алгоритм, який дозволяє за рахунок використання отриманих даних скласти загальну картину застосування визначеного поняття у різних наукових галузях.

Отримані результати дозволяють побудувати нові масиви інформації за рахунок використання наукометричної інформації, що надається ресурсами відкритого доступу: енциклопедичними, наукометричними та архівом препринтів.

Четвертий розділ дисертаційної роботи присвячено опису можливостей реалізації запропонованих технологій для: мережі співавторів,

що побудована для таких базових понять, як text mining, language processing, information retrieval, complex network; мережі предметних областей, яка побудована для понять Physical optics, Scientometric; Вікі-індексу автора на базі енциклопедичного ресурсу Wikipedia; мережі предметних областей на базі ресурсу препрінтів arXiv для поняття ‘cavitation’ та ‘Ukraine.’

Також наведено результати порівняльного аналізу запропонованого показника із притаманним в наукометрії для різних наукометричних систем індексами.

У висновках наводяться основні наукові та практичні результати.

У додатках наведено результати аналізу систем наукової інформації з урахуванням запропонованих критеріїв, лістинги програми реалізації алгоритму побудови онтології мовою Java на базі он-лайн енциклопедії Вікіпедія, словники предметних областей для ресурсу препрінтів arXiv, які містять переліки наукових напрямків, що конкретизують відповідну предметну область.

Отже поставлені наукові завдання в повному обсязі вирішено та наведено в дисертаційній роботі. Вищенаведене дозволяє зробити висновок про відповідність назви дисертації її змісту.

Відповідність змісту автореферату основним положенням дисертації. Зміст автореферату повністю відображає основні результати досліджень, які подані в дисертації. В авторефераті в повній мірі викладені усі наукові положення та результати з достатньою детальністю.

Недоліки та зауваження.

1. Постановку завдання наукового дослідження зроблено в узагальненій описовій формі, що дещо ускладнює розуміння її та оцінку результатів, що були отримані автором в ході наукового дослідження.

2. У розділі 1 дисертаційної роботи, автор дає означення інформаційної системи, проте під час опису інформації, яку вони обробляють, користується терміном «інформаційно-аналітична система», який, на мою думку, доцільно було б розкрити у тексті дисертації та включити даний клас інформаційних систем до класифікатора, який наведено автором у цьому ж розділі.

3. На мій погляд наведений перелік вимог, яких мають дотримуватися інформаційні системи для організації корисної взаємодії з користувачем (стор. 42) є досить узагальненим та потребує розширення.

4. Запропонована в роботі описова модель для зображення наукової інформації (п. 1.6), має припущення щодо однакової ваги всіх критеріїв, які застосовуються для оцінки систем аналізу наукової інформації. Це, на мою думку, дещо спрощує модель та потребує її подальшого удосконалення.

5. У розділі 2 дисертації, на рис. 2.1 зображено чотирьохрівневу описову модель ресурсу наукової інформації без конкретного позначення її рівнів, що було б цілком доцільно для її повного розуміння.

6. Для побудови мережі предметних областей у дисертації (розділ 3) запропоновано метод зондування опорної мережі, який реалізовано у вигляді оригінального алгоритму та проілюстровано на прикладі моделі випадкового графу Ердеша-Рені та випадкової безмасштабної мережі Барабаші-Альберт. Проте відповідної інтерпретації отриманих результатів в роботі не наведено.

7. Запропоновані в дисертаційній роботі методи побудови мереж предметних областей на базі різних ресурсів (Wikipedia, Google Scholar Citations, ресурсу препрінтів arXiv) реалізовано у вигляді алгоритмів, проте оцінка їхньої обчислювальної складності, в роботі не наведена.

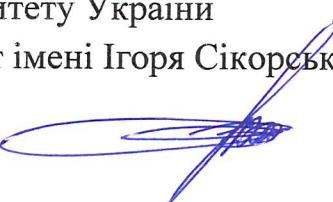
Проте, зазначені недоліки не знижують ступінь наукової новизни та практичного значення одержаних в дисертації наукових результатів і, відповідно, позитивну оцінку роботи цілому.

Висновок. Дисертаційна робота Андрушенко В.Б. є кваліфікаційною науковою працею, яка містить нові наукові результати, що в сукупності вирішують актуальне наукове завдання розробки інформаційних технологій для проведення наукометричного аналізу на основі моніторингу ресурсів мережі Інтернет, формування масивів інформації, розробки відповідних моделей і алгоритмів реалізації додаткових можливостей для існуючих наукометричних ресурсів та ресурсів наукової інформації відкритого доступу. Дисертація відповідає вимогам п.п. 9, 11, 12 Порядку присудження наукових ступенів, затвердженого постановою Кабінету Міністрів України від 24.07.2013 № 567, а її автор, Андрушенко Валентина Борисівна, заслуговує присудження наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології.

Завідувач спеціальної кафедри № 5
Інституту спеціального зв'язку та захисту інформації
Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського»

доктор технічних наук, доцент

«19» березня 2019 року



I.Y.O. Субач

Заступник начальника інституту
(з наукової роботи)
кандидат технічних наук, доцент



C.M. Конюшок