

АНОТАЦІЯ

Главчева Ю. М. Методи та інформаційна технологія інформетрії в академічному інформаційному середовищі. Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії (PhD) за спеціальністю 122 – Комп’ютерні науки – Національний технічний університет “Харківський політехнічний інститут”, Харків, 2020.

Об’єктом дослідження є процес оцінювання наукової діяльності.

Предметом дослідження є моделі, методи та інформаційна технологія підвищення якості процесу оцінки наукової діяльності.

У вступі обґрунтовано актуальність досліджуваної теми. Описано зв’язок роботи з науковими програмами, планами та темами, наведено наукову новизну, а також, сформульовано практичне значення отриманих результатів.

В першому розділі здійснено аналітичний огляд підходів до реалізації процесу оцінки наукової діяльності. Виконано аналіз інформаційно-аналітичних пошукових систем, виділено авторитетні джерела наукометричних даних, проаналізовано основні наукометричні показники та метрики, які застосовуються на практиці. Визначено складові наукометричних систем та фактори, які впливають на їх якість. Під оцінкою наукової діяльності розуміємо дослідження якості наукового процесу, а саме якісного проведення усіх етапів наукової роботи, у тому числі якості наукових публікацій. Якість та оригінальність наукового дослідження може бути забезпечено унеможливленням академічного шахрайства, тобто виявленням потенційного академічного плагіату. Оригінальним текстом є текст, який створений автором самостійно, у якому всі використані текстові запозичення мають посилання на першоджерело. Тому дослідження включає завдання зі створення інформаційної технології визначення авторського стилю для ідентифікації автора та методу виділення авторської ідеї,

пошук документів, ідеї в яких є семантично подібні. На основі аналізу складових процесу оцінки наукової діяльності сформовано основні напрямки досліджень та визначено основні задачі дисертаційної роботи.

В другому розділі розглянуто питання та описана розроблена інформаційна технологія визначення авторського стилю в академічних текстах українською та російською мовами. Дослідниками світу виділяються різні групи стилOMETричних властивостей. Але не всі властивості однаково ефективні для будь-якої мови та будь-яких текстів. У роботі групу стилOMETричних властивостей було застосовано до академічних текстів українською та російською мовами. Сформовано модель представлення тексту у вигляді стилOMETричних наборів ознак. На базі цієї моделі виділено стилOMETричні властивості, за допомогою яких можлива ідентифікація автора тексту. Запропоновано використання методів контекстного моделювання та машинного навчання для проведення класифікації текстів на базі стилOMETричних властивостей за визначеними класами (за авторами).

В третьому розділі розглянуто питання визначення авторської ідеї академічного тексту українською мовою для визначення замаскованого інтелектуального академічного плагіату. Адаптовано графовий метод (TextRank) до особливостей наукового тексту для вилучення основної ідеї тексту, тобто формування реферату. Розроблено метод визначення авторської ідеї на основі аналізу анафоричних зв'язків та адаптованого методу TextRank. Для опису семантичних зв'язків між реченнями використано апарат алгебри скінченних предикатів та предикатних операцій, метод компараторної ідентифікації, побудовано логічну мережу.

Четвертий розділ присвячено розгляду методів побудови рейтингів академічних установ. Розглянуто методи оцінювання наукової діяльності академічних установ, та чинники, що впливають на якість результату. Досліджено особливості публікаційної поведінки авторів. Визначено, що

пріоритетне використання даних з публікаційної активності може призводити до маніпулювання цими даними. Проведено аналіз характеристик переліку використаних джерел. Розроблено метод до визначення якості переліку використаних джерел, як одного з якісних характеристик наукового документу. Аналіз національного рейтингу академічних установ Scopus показав певні недоліки. Запропоновано методику ранжування академічних установ з використанням даних наукометричних систем Scopus та Web of Science CC, яка враховує особливості оцінюваних об'єктів та обмеження бібліометричних методів.

У п'ятому розділі виконано практичну реалізацію запропонованих інформаційної технології та методів. Показано використання коефіцієнтів різноманітності та статистичних параметрів тексту, лексичні та символльні властивості для ідентифікації авторського стилю написання тексту українською та російською мовами. Застосовано метод контекстного моделювання Prediction by Partial Matching та методи машинного навчання Naive Bayes, Naive Bayes Multinomial, Support Vector Machine, Decision Trees для вирішення задачі класифікації, а саме за визначеними авторів. Апробовано використання розробленої методики з виділення основної ідеї тексту та з порівняння ідей двох документів.

Таким чином, у дисертаційній роботі вирішується науково-прикладна задача з забезпечення підвищення якості процесу оцінки наукової діяльності. Її рішення реалізується шляхом розробки інформаційної технології, методів, методики та програмних додатків, що створюють передумови для підвищення якості результатів оцінки наукової діяльності за рахунок сприяння академічній доброчесності у академічному середовищі шляхом використання програмних засобів, забезпечення достовірності наукометричних показників та справедливих підходів при ранжуванні оцінюваних об'єктів.

В роботі *вперше* отримано такі результати: *вперше* розроблено метод ідентифікації авторської ідеї академічного тексту, на основі інтелектуального аналізу даних, який відрізняється від існуючих новизною, та дозволяє визначити ознаки прихованого (інтелектуального) академічного плагіату в наукових роботах; *вперше* розроблено метод оцінки якості переліку використаних джерел, як одного з критеріїв оцінки якості статті, що дозволило акцентувати увагу рецензента та запобігти можливій маніпулятивній поведінці авторів.

Отримала подальший розвиток інформаційна технологія визначення авторського стилю україномовних та російськомовних одноосібних академічних публікацій. Ця технологія заснована на методах машинного навчання і лінгвістичних технологіях, що дозволило створити єдину технологію визначення авторства.

Удосконалено методику формування рейтингу академічних установ на основі показників інформетрії, яка враховує особливості досліджуваних об'єктів та обмеження бібліометричних методів.

Дисертація є частиною науково-дослідних робіт, виконаних на кафедрі Інтелектуальних комп'ютерних систем «Розробка моделей та методів ідентифікації процесів екстракції знань у слабоструктурованих масивах інформації» (2017/2018 р., № ДР 0117U004726).

Розроблені інформаційна технологія, методи, методика та програмний додаток *впроваджені*:

- в ТОВ «Видавнича служба УРАН» Асоціації користувачів Української науково-освітньої телекомунікаційної мережі «УРАН» реалізується метод ідентифікації авторської ідеї академічного тексту, на основі інтелектуального аналізу даних;

- в процес обробки текстових даних при вирішенні практичних завдань «Фірма SOLTI» упроваджено інформаційну технологію обробки природньої мови та визначення письмового стилю для ідентифікації автора;

– в роботу Наукової бібліотеки Національного юридичного університету ім. Ярослава Мудрого;

– у науково-технічній бібліотеці Національного технічного університету «Харківський політехнічний інститут» з метою забезпечення професійного виконання наукометричної експертної оцінки та задля якісного та швидкого задоволення методичних та інформаційних потреб користувачів;

– в навчальний процес кафедри інформаційних ресурсів та комунікацій освітньої установи Білоруського державного університету культури та мистецтв (м. Мінськ, Білорусь);

– в навчальний процес кафедри інтелектуальних комп'ютерних систем факультету соціально-гуманітарних технологій Національного технічного університету «Харківський політехнічний інститут» (м. Харків, Україна).

Ключові слова: оцінка наукової діяльності, наукометрія, академічний плагіат, академічна доброчесність, авторський стиль, визначення авторської ідеї, академічні рейтинги.

Список публікацій здобувача.

Наукові праці, в яких опубліковано основні наукові результати:

1. Glavcheva Y., Kanishcheva O., Glavchev M. Evaluating the quality of research activities: Investigating into the uniqueness. *Scientific and Technical Libraries*. 2018. №10. P. 5-21. **(Web of Science CC)**.
2. Главчева Ю. М., Каніщева О. В. Ідентифікація текстового плагіату в академічних документах. *Системи управління, навігації та зв'язку*. 2018. №6(52), С. 83-86. Режим доступу: <http://journals.pntu.edu.ua/sunz/article/view/1331> (дата звернення: 29.09.2020).

3. Главчева Ю. М., Главчев М. І., Каніщева О. В., Кучук Г. А. Розробка підходу для ранжування академічних установ за показниками наукової діяльності. *Сучасні інформаційні системи*. 2019. Т. 3. № 1. С. 63-71. Режим доступу: <http://ais.khpi.edu.ua/article/view/2522-9052.2019.1.11> (дата звернення: 29.09.2020).
4. Главчева Ю. М., Каніщева О. В., Борисова Н. В. Огляд методів та технологій інформетрії (Survey of informetry methods and technologies). *Кібернетика і Системний Аналіз*. 2019. №3. С. 182–193. Режим доступу: <https://link.springer.com/article/10.1007%2Fs10559-019-00158-z> (дата звернення: 29.09.2020) (Hlavcheva, Y. M., Kanishcheva, O. V., & Borysova, N. V. A survey of informetric methods and technologies. *Cybernetics and Systems Analysis*. 2019. № 55(3), 503-513. doi:10.1007/s10559-019-00158-z) (**Scopus, Q2**).
5. Главчева Ю. Главчев М. До питання оцінки якості науки. *Системи управління, навігації та зв'язку*. 2020. Т. 1. №. 59. С. 63-67. Режим доступу: <http://journals.nupr.edu.ua/sunz/article/view/1784> (дата звернення: 29.09.2020).
6. Hlavcheva Y. M. Using Statistical Text Parameters to Determine Author's Scientific Documents Written Style. *Polish Journal of Science*. 2020. VOL. 1. №. 32. P. 57-60.
7. Hlavcheva Y., Kanishcheva O., Vovk M. Identifying Semantic Relations Between Sentences by Solving an Anaphora. *Information Processing Systems*. 2020. № 3(162). – P. 36-43. doi:10.30748/soi.2020.162.04.

Опубліковані праці апробаційного характеру:

8. Каніщева О. В., Главчева Ю. Н., Висоцька В. А. Визначення стилю автора для виявлення плагіату в академічному середовищі. *Системний аналіз та інформаційні технології* : матеріали 19-ї Міжнар. наук.-технічн. конф. SAIT

2017, 22-25 травня 2017. Київ : ННК "ІПСА" НТУУ "КПІ" ім. Ігоря Сікорського, 2017. С. 78-79.

9. Главчев М. І., Главчева Ю. Н., Каніщева О. В. Забезпечення достовірних показників інформаційно-аналітичних систем. *Проблеми інформатики та моделювання* : тези 17-ї Міжнар. наук.-технічн. конф., 11-15 вересня 2017. Харків : НТУ "ХПІ", 2017. С. 19-20.
10. Главчева Ю. М., Каніщева О. В., Главчев М. І. Визначення авторського стилю в академічних текстах: емпіричні дослідження. *Інформатика, управління та штучний інтелект* : матеріали 4-ї міжнар. наук.-техн. конф. студентів, магістрів та аспірантів, 21-23 листопада 2017 р. Харків : НТУ "ХПІ", 2017. С. 33. Режим доступу: <http://repository.kpi.kharkov.ua/handle/KhPI-Press/33009> (дата звернення: 29.09.2020).
11. Главчева Ю. М. До питання оцінки наукової діяльності. *Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління* : тези доповідей VIII міжнародної науково-технічної конференції, 26-27 квітня 2018 р. Харків, 2018. С. 14.
12. Главчева Ю. М., Главчев М. І. Удосконалення програмно-технічних засобів задля підтвердження оригінальності академічних текстів. *Інформаційні технології: наука, техніка, технологія, освіта, здоров'я* : тези доповідей XXVI міжнародної науково-практичної конференції, 16-18 травня 2018 р. Харків, 2018. С. 166.
13. Главчева Ю. М. Метод порівняння академічних установ на основі визначення долі публікацій, що мають цитування вище середнього. *Модель бібліотеки XXI століття* : матеріали міжнар. наук.-практ. інтернет-конф., 15 червня 2018 р. Нац. юрид. ун-т ім. Ярослава Мудрого, Наук. б-ка. Електрон. текст. дані. Харків, 2018.

- 14.Vysotska V., Kanishcheva O., Hlavcheva Y. Authorship Identification of the Scientific Text in Ukrainian with Using the Lingvometry Methods. *2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, Lviv, 2018, pp. 34-38. doi: 10.1109/STC-CSIT.2018.8526735 **(Scopus)**.
- 15.Главчева Ю. М., Главчев М. І., Каніщева О. В. Розробка універсального методу оцінки наукової діяльності. *Проблеми інформатики та моделювання* : тези 18-ї Міжнар. наук.-технічн. конф., 15-19 вересня 2018. Харків : НТУ "ХПІ", 2018. С. 4.
- 16.Главчева Ю. М. Аналіз технологій виявлення текстового плагіату. *Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління* : тези доповідей IX міжнародної науково-технічної конференції, 11-12 квітня 2019 р. Харків, 2019. С. 18.
- 17.Bobicev V., Hlavcheva Y., Kanishcheva O., Lazu V. Authorship Attribution in Scientific Publications. *Proceedings of Corpora-2019 conference*. Saint-Petersburg, Russia, pp. 174-181.
- 18.Bobicev V., Hlavcheva Y., Kanishcheva O. Can Computers Catch the Authors Style? *The Fifth Conference of Mathematical Society of the Republic of Moldova*. 2019. С. 293.
- 19.Главчева Ю. М. Виявлення подібностей у тексті: зовнішні та внутрішні методи. *Проблеми інформатизації* : тези доповідей VII міжнародної науково-технічної конференції, 13-15 листопада 2019 р. 2019. С. 78.
- 20.Главчева Ю. М. Розробка методу виявлення інтелектуального плагіату. *Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління* : тези доповідей десятої міжнародної науково-технічної конференції, 9-10 квітня 2020 р. 2020. Том 2: секції 3, 4. С. 41. Режим доступу: <http://repository.kpi.kharkov.ua/bitstream/KhPI->

Press/45561/1/Conference_NTU_KhPI_2020_Suchasni_napriamy_rozvytku_I
KT_T_2.pdf (дата звернення: 29.09.2020).

- 21.Главчева Ю. М., Каніщева О. В. Актуальні напрямки розвитку технологій виявлення текстового академічного плагіату. *Інформаційні технології: наука, техніка, технологія, освіта, здоров'я* : тези доповідей XXVIII міжнародної науково-практичної конференції, 28-29 жовтня 2020 р. Харків, 2020.
- 22.Hlavcheva Y., Kanishcheva O. The Reference Analysis as a Quality Characteristic of a Scientific Article. *ICTERI*. — 2020. (**Scopus**).

Опубліковані праці які додатково відображають наукові результати:

- 23.Главчева Ю. М. Куратор змісту / Главчева Ю. М., Кухаренко В. М., Рибалко О. В.. За ред.. В. М. Кухаренка – ХПІ, Харків:, 2016. – 157 с.
- 24.Методичні рекомендації з редакційної підготовки наукових журналів : для редакційних колегій наукових видань НТУ "ХПІ" / уклад.: Ю. М. Главчева, Г. Б. Обухова, М. І. Безменов ; Нац. техн. ун-т "Харків. політехн. ін-т". – Харків : НТУ "ХПІ", 2018. – 39 с.
- 25.Главчева Ю. М. Академічна доброчесність, як запорука підвищення якості / Ю. М. Главчева // Академічна доброчесність: виклики сучасності : збірник наукових есе учасників наукового стажування для освітян (Республіка Польща, Варшава, 1–13.10.2018). — Варшава, 2018. – С. 24-27.