***The Current State of Automated Analysis of Scientific Publication Activity Using Digital Identifier Data***

|  |
| --- |
| Dmytro Miach, Yana Romashkevych, Mariia SoldatovaFaculty of Informatic and Computer EngineeringIgor Sikorsky Kyiv Polytechnical Institute Kyiv, Ukrained.miach@kpi.ua, y.romashkevych@gmail.com, benten1093@gmail.com   |
| ***Abstract.* Monitoring the publication output of academic staff is crucial for internal performance evaluation and for shaping an institution’s profile in external rankings. Automated analyses leveraging digital identifier data aim to streamline and accelerate this process but are subject to specific vulnerabilities, which are examined in this article.** ***Keywords*: *DOI, publication activity, bibliographic description, information system.*** |

***Cучасний стан автоматизованого аналізу публікаційної активності науковців на основі даних цифрового ідентифікатора***

М’яч Дмитро Олександрович, Ромашкевич Яна Олександрівна, Солдатова Марія Олександрівна

КПІ ім. Ігоря Сікорського

м. Київ, Україна

d.miach@kpi.ua, y.romashkevych@gmail.com, benten1093@gmail.com

*Анотація.* Публікаційна активність викладачів та науковців є важливим критерієм в контексті проведення внутрішнього оцінювання (рейтингування) роботи науково-педагогічних працівників, а також для формування портрету закладу вищої освіти (ЗВО) в зовнішніх рейтингах. Автоматизований аналіз публікаційної активності на основі даних цифрового ідентифікатора покликаний спростити та пришвидшити даний процес, проте має окремі вразливості, які будуть розглянуті в контексті даної публікації.

# *Ключові слова: DOI, публікаційна активність, бібліографічний опис, інформаційна система.*

Вступ

 Публікаційна активність науково-педагогічного працівника (далі - НПП) відіграє важливу роль у виконанні ним Кадрових вимоги щодо освітньої діяльності за рівнем вищої освіти відповідно до Ліцензійних умов провадження освітньої діяльності [1], а також в межах КПІ ім. Ігоря Сікорського – в процесі проведення конкурсного відбору або обрання за конкурсом при заміщенні вакантних посад науково-педагогічних працівників. [2].

 Впровадження автоматизованого аналізу публікаційної активності науковців має забезпечити збір інформації щодо публікацій і забезпечення уніфікації в підсистемі рейтингування для подальшого використання.

# Аналіз існуючого стану збору даних

З 2017 року Науково-технічна бібліотека ім. Г. І. Денисенка (далі – НТБ) здійснює моніторинг та аналіз публікаційної активності НПП та вчених, які афілійовані з
КПІ ім. Ігоря Сікорського. Базою для моніторингу є наукометричні бази даних Scopus та Web of Science Core Collection.

За результатами моніторингу та аналізу:

* укладаються переліки працівників, публікації яких оприлюднені у виданнях, що входять до міжнародних наукометричних баз даних Scopus та/або Web of Science Core Collection та відповідають критеріям преміювання;
* укладають переліки нових публікацій вчених
КПІ ім. Ігоря Сікорського [3];
* з 2020 року ведеться облік усіх публікацій (з усіма показниками впливовості);
* готуються додатки на запит НДЧ для подальшого подання до ЄДЕБО (атестація за науковими напрямами, держзамовлення магістрів тощо).

З 2020 року в межах автоматизованої інформаційної системи «Електронний Кампус» реалізовано підсистему «Рейтинг НПП», в якій відповідно до Норм бального оцінювання діяльності науково-педагогічних працівників здійснювався збір інформації щодо виконання НПП зазначених критеріїв із їх подальшим оцінюванням [4].

Відповідно до вказаних Норм, проводиться оцінювання публікацій НПП відповідно до критеріїв щодо категорій видань, а також щодо участі в співавторах здобувачів вищої освіти та/або іноземних учених [5].

В існуючій системі внесення та перевірка внесених даних відбувається в ручному режимі шляхом залучення фахівців НТБ у ролі «відповідального за доповнення інформації» та «адміністратору пункту» [6].

 

Рис. 1. Кількість записів за напрямками діяльності
з науково-інноваційної роботи за 2023-2024 н.р.

 Відповідно до даних, наявних в підсистемі «Рейтинг НПП» за 2023-2024 н. р. публікації складають 22,8% записів від всіх записів внесених за напрямком «Науково-інноваційна робота» (рис.1).

Таблиця 1

Відомості щодо внесених DOI статей

|  |  |  |
| --- | --- | --- |
| **Тип публікації (статті)** | **Всього записів** | **Наявність DOI** |
| – стаття у журналах, що індексуються базами Scopus та/або Web of Science\*, віднесених до Q1 | 198 | 173 |
| – стаття у журналах, що індексуються базами Scopus та/або Web of Science\*, віднесених до Q2 | 237 | 218 |
| – стаття у журналах, що індексуються базами Scopus та/або Web of Science\*, віднесених до Q3 | 600 | 525 |
| – стаття у журналах, що індексуються базами Scopus та/або Web of Science\*, віднесених до Q4 | 628 | 492 |
| – стаття у журналах, що індексуються базами Scopus та/або Web of Science\*, та не мають квартилю | 103 | 70 |
| – стаття у зарубіжних періодичних наукових виданнях країн ОЕСР, що не індексуються базами Scopus та/або Web of Science | 224 | 107 |
| – стаття у фахових журналах категорії Б, що не індексуються базами Scopus та/або Web of Science | 2356 | 1501 |
| – стаття у фахових журналах категорії Б, що не індексуються базами Scopus та/або Web of Science (англійською) | 933 | 647 |

 Вимога щодо внесення DOI (digital object identifier) для публікацій була необов’язковою, але як видно з табл. 1, в якій наведено інформацію щодо внесених даних про DOI в описі публікацій, то більше ніж в 67% статей, що індексуються базами Scopus та/або Web of Science, вказано даний ідентифікатор, у видань категорії Б це значення коливається від 63 до 69%

# Використання зовнішнього ідентифікатору публікацій для аналізу

 DOI як унікальний цифровий ідентифікатор об’єкта (згідно стандарту ISO 26324) з одного боку, а також інструмент по скороченню посилань з іншого боку, виступає багатофункціональним інструментом для уніфікації інформації щодо внесених публікацій [7].

 Експериментальне очищення даних від дублікатів щодо публікацій, що містить DOI без їх додаткової верифікації, показало, що із 3765 записів за текстовим описом унікальними виявились 3600 записів, а за DOI – 2760. Даний показник доводить, що ручне внесення інформації різними особами (надавачами) призводить до виникнення більшої кількості дублюючих записів, які не піддаються автоматичному виявленню.

 Аналіз публікація на дублікати важливий для того, щоб уникнути спотворення інформації при формуванні подальших звітів в контексті, який буде відділений від сутності працівника в контекст публікацій.

 За допомогою API від Crossref із використанням DOI можна отримати інформацію щодо метаданих для кожної із публікацій, які передаються видавцем до кожного із ідентифікаторів [8].

 Подальше очищення даних шляхом перевірки коректності DOI та наявності активних метаданих щодо публікації дозволило відсіяти ще 11% записів.

 Безпосередньо для наповнення бази публікаційної активності із метою подальшого аналізу можна використовувати методи Crossref REST API, які повертають метадані щодо публікацій у вигляді JSON-файлів.

 В структурі об’єкту, що повертається на запит по конкретному DOI, особливо важливими для автоматизування системи можна виділити наступні поля:

* PublishedDate –дозволяє визначити чи відноситься публікація до вказаного періоду рейтингування (оцінювання);
* Author\_1\_Orcid, Author\_2\_Orcid тощо – дозволяє автоматизовано визначити особу автора в системі;
* Author\_1\_Affiliation, Author\_2\_Affiliation тощо – дозволяє визначити чи вказано місце місце роботи автора, в іншому випадку – чи вказано саме
КПІ ім. Ігоря Сікорського, якщо науковець подає її в межах Університету.

Не менш критичним для автоматизування процесу обробки інформації в межах системи та зменшення навантаження на працівників, призначених на ролі «відповідального за доповнення інформації» та «адміністратору пункту» є наповнення довідника Університету даними щодо ORCID профілів науковців.

Важливим є те, щоб ORCID профіль науковця був актуальним і використовувався автором при подачі публікацій.

В такому випадку роль НТБ від постійного збору інформації переходить до верифікація та вдосконалення метаданих.

В контексті бази публікаційної активності бібліотекарі можуть і повинні: верифікувати надіслані авторами DOI, виправляти та уточнювати метадані (наприклад, уніфікувати афіліації), контролювати коректність записів у базі, щоб дані були точними й відповідали стандартам.

У табл. 2 наведено показники проаналізованих метаданих публікацій, що вже наявні в підсистемі «Рейтинг НПП» та були внесені з вказанням DOI із відповідними полями, які необхідні для формування бібліографічного опису публікацій із подальшим збереження в базу публікаційної активності.

Таблиця 2

Статистика щодо полів метаданих публікацій

|  |  |  |  |
| --- | --- | --- | --- |
| **Поле JSON** | **К-сть** | **%** | **Відповідник** |
| DOI | 2451 | 100 |  |
| Status | 2451 | 100 | Актуальність ідентифікатора |
| Title | 2398 | 97,84 | Назва публікації |
| Publisher | 2448 | 99,88 | Видавець |
| PublishedDate | 2437 | 99,43 | Дата видання |
| Type | 2448 | 99,88 | Тип публікації |
| Author\_1\_Given | 2091 | 85,31 | Автор 1, ім'я |
| Author\_1\_Family | 2101 | 85,72 | Автор 1, прізвище |
| Author\_1\_Orcid | 1149 | 46,88 | Автор 1, ORCID |
| Author\_1\_Affiliation | 331 | 13,50 | Автор 1, афіляція |
| Author\_2\_Given | 2134 | 87,07 | Автор 2, ім'я |
| Author\_2\_Family | 2143 | 87,43 | Автор 2, прізвище |
| Author\_2\_Orcid | 1151 | 46,96 | Автор 2, ORCID |
| Author\_2\_Affiliation | 317 | 12,93 | Автор 2, афіляція |
| Author\_3\_Given | 1426 | 58,18 | Автор 3, ім'я |
| Author\_3\_Family | 1433 | 58,47 | Автор 3, прізвище |
| Author\_3\_Orcid | 722 | 29,46 | Автор 3, ORCID |
| Author\_3\_Affiliation | 266 | 10,85 | Автор 3, афіляція |
| Author\_4\_Given | 830 | 33,86 | Автор 4, ім'я |
| Author\_4\_Family | 835 | 34,07 | Автор 4, прізвище |
| Author\_4\_Orcid | 411 | 16,77 | Автор 4, ORCID |
| Author\_4\_Affiliation | 214 | 8,73 | Автор 4, афіляція |
| Author\_5\_Given | 658 | 26,85 | Автор 5, ім'я |
| Author\_5\_Family | 661 | 26,97 | Автор 5, прізвище |
| Author\_5\_Orcid | 326 | 13,30 | Автор 5, ORCID |
| Author\_5\_Affiliation | 152 | 6,20 | Автор 5, афіляція |

 Відштовхуючись від даних табл. 2 варто зауважити, що необов’язковість полів при заповненні метаданих публікацій призводить до відсутності в половини авторів вказаного ORCID. Проте, зважаючи на об’єми даних, які підлягають обробці, вказаний відсоток дозволяє частково автоматизувати процес.

# Висновки

Використання цифрового ідентифікатора DOI має важливе значення для покращення процесу збору, уніфікації та аналізу публікаційної діяльності науковця, проте потребує залучення додаткових інформаційних та людських ресурсів для повноти отриманої інформації.

##### Література

1. Про затвердження Ліцензійних умов провадження освітньої діяльності: Постанова КМУ від 30.12.2015 № 1187: станом на 4 січ. 2024 р.
URL: [https://zakon.rada.gov.ua/laws/show/1187-2015-п#Text](https://zakon.rada.gov.ua/laws/show/1187-2015-%D0%BF#Text) (дата звернення: 12.05.2025)
2. Про затвердження Порядку проведення конкурсного відбору або обрання за конкурсом при заміщенні вакантних посад науково-педагогічних працівників та укладання з ними трудових договорів (контрактів): Наказ КПІ ім. Ігоря Сікорського від 24.09.2021 р. № НУ/201/2021. URL: https://osvita.kpi.ua/sites/default/files/downloads/Наказ\_№НУ\_201\_2021\_від\_24\_09\_2021.pdf .
3. Оцінка | Науково-технічна бiблiотека ім. Г. І. Денисенка. URL: https://www.library.kpi.ua/research/otsinka/#2 (дата звернення: 09.05.2025).
4. М’яч Д. О. Підсистема «Рейтинг НПП» АІС «Електронний кампус»: індивідуальний дослідницький проєкт… бакалавра: 126 Інформаційні системи та технології / М’яч Дмитро Олександрович. — Київ, 2022. — 64 с. URL: <https://ela.kpi.ua/handle/123456789/57792> (дата звернення 09.05.2025)
5. Про затвердження Положення про рейтингування науково-педагогічних працівників КПІ ім. Ігоря
Сікорського та норм бального оцінювання діяльності науково-педагогічних працівників: Наказ від 25.12.2023 р. № НОН/386/2023
6. Розподіл ролей користувачів системи рейтингування науково-педагогічних працівників / М. О. Солдатова та ін. Телекомунікаційні та інформаційні технології. 2024. Т. 82, № 1. С. 106–113. URL: <https://doi.org/10.31673/2412-4338.2024.019901> (дата звернення: 07.05.2025).
7. DOI® Resolution Documentation. URL: https://doi.org/the-identifier/resources/factsheets/doi-resolution-documentation (date of access: 08.05.2025).
8. Crossref REST API. Swagger UI. URL: https://api.crossref.org/swagger-ui/index.html (date of access: 08.05.2025).