

УДК 004.41:378.1

О. Ю. Мельников

СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ВИЗНАЧЕННЯ НАЛЕЖНОСТІ ОБ'ЄКТІВ НАУКОВО-ОСВІТНЬОЇ ДІЯЛЬНОСТІ НАУКОВО-ОСВІТНІМ ГАЛУЗЯМ І СПЕЦІАЛЬНОСТЯМ

Донбаська державна машинобудівна академія, Краматорськ

Анотація. Розглянуто задачу віднесення об'єкта науково-освітньої діяльності (викладача, кафедри, студентської групи тощо) до певної науково-освітньої галузі знань або спеціальності. Зазначено, що розв'язання цієї задачі важливо при акредитації освітніх програм, при виборі рецензента для наукової статті або студентської конкурсної наукової праці, іноді – при виборі керівника кваліфікаційної роботи тощо. Було поставлено та розв'язано задачу створення інформаційної системи (системи підтримки прийняття рішень), яка сприяла б формуванню показників діяльності та проводила перевірку відповідності цих показників галузям освітньої діяльності або науковим спеціальностям. Перелік результатів діяльності об'єкта може при цьому міститися у наукових публікаціях, у ключових словах, якими науковець описує свої наукові інтереси, або в темах кваліфікаційних студентських робіт. Основними інформаційними ресурсами є база категоризованих наукових публікацій із системи Dimensions та інформація про науковців у системі Google Scholar. Використовуються системи класифікації ANZSRC-2008, ANZSRC-2020, ISCED-F та стандарт, прийнятий в Україні. Описано приклади роботи розробленої системи, розрахунок показників за різними формулами (абсолютні значення числа знайдених входжень; частка кожного значення стосовно суми всіх показників досліджуваного об'єкта; частота входження в тематичну колекцію по кожній галузі або спеціальності; частка кожної частоти входження до суми всіх часток досліджуваного об'єкта); аналіз результатів розрахунків як у кожній позиції (парі ключових слів чи наукової публікації) досліджуваного об'єкта, і загального показника (який, своєю чергою, можна розрахувати або як середнє арифметичне, або як суму значень); «скорочення» результатів («відсікання хвостів») або за мінімальним рівнем, або за максимальним числом позицій, з можливістю нормалізації підсумку (приведення суми до одиниці); переклад розподілу за галузями та спеціальностями ANZSRC до таблиці ISCED-F або до стандарту, прийнятого в Україні; розрахунок показників по всій групі об'єктів та ранжування об'єктів у цій групі за обраною галуззю чи спеціальністю.

Ключові слова: науково-освітня діяльність, категоризація, ранжування, Google Scholar, Dimensions, ANZSRC, ISCED-F, галузь знань, спеціальність.

Abstract. The task of assigning the object of scientific and educational activity (teacher, department, student group, etc.) to a certain scientific and educational field of knowledge or specialty is considered. It is noted that solving this problem is important during the accreditation of educational programs, when choosing a reviewer for a scientific article or a student competitive scientific paper, sometimes - when choosing a supervisor for a qualification work, etc. The task of creating an information system (decision-making support system) was set and solved, which would contribute to the formation of performance indicators and check the compliance of these indicators with the fields of educational activity or scientific specialties. At the same time, the list of results of the object's activity can be contained in scientific publications, in keywords used by the scientist to describe his scientific interests, or in the topics of qualifying student papers. The main information resources are the database of categorized scientific publications from the Dimensions system and information about scientists in the Google Scholar system. The classification systems ANZSRC-2008, ANZSRC-2020, ISCED-F and the standard adopted in Ukraine are used. Examples of the work of the developed system are described, the calculation of indicators according to various formulas (absolute values of the number of occurrences found; the share of each value in relation to the sum of all indicators of the object under study; the frequency of entry into the thematic collection for each field or specialty; the share of each frequency of entry to the sum of all particles of the object under study); analysis of the results of calculations both in each position (keyword pair or scientific publication) of the researched object, and the general indicator (which, in turn, can be calculated either as an arithmetic average or as a sum of values); "reduction" of the results ("cutting off the tails") either by the minimum level or by the maximum number of positions, with the possibility of normalizing the total (bringing the sum to one); translation of the distribution by branches and specialties of the ANZSRC to the ISCED-F table or to the standard adopted in Ukraine; calculation of indicators for the entire group of objects and ranking of objects in this group by the chosen industry or specialty.

Key words: scientific and educational activity, categorization, ranking, Google Scholar, Dimensions, ANZSRC, ISCED-F, field of knowledge, specialty.

DOI: <https://doi.org/10.31649/1999-9941-2023-57-2-109-116>.

Вступ

Задача віднесення науково-педагогічного співробітника до певної науково-освітньої галузі знань або спеціальності традиційно є однією з важливих та одночасно складних завдань. Відповідність викладача (наукового співробітника) тієї або іншої галузі важливо при акредитації освітніх програм, при виборі рецензента для наукової статті або студентської конкурсної наукової праці, іноді – при виборі керівника кваліфікаційної роботи тощо.

Актуальність та мета

Як правило, як критерій відбору використовуються [1]:

- базова освіта;
- наукова спеціальність (по якій було здійснено захист дисертації);
- тематика наукових публікацій за останні п'ять років.

Такий підхід містить небезпеку суб'єктивності та категоричності (звичай потрібно не визначати єдину відповідну спеціальність, а виділити перелік суміжних галузей). Очевидно, що створення моделі та системи підтримки прийняття підсилить об'єктивність цієї діяльності.

Задачі

Сформулюємо задачу таким чином. Є об'єкт науково-освітньої діяльності (викладач, кафедра, студентська група), який необхідно співвіднести з категоріями наукових спеціальностей або освітніх галузей – визначити ступінь відповідності діяльності об'єкта кожної категорії.

Діяльність об'єкта може виражатися:

- у наукових публікаціях;
- у ключових словах, якими науковець описує свої наукові інтереси;
- у темах кваліфікаційних студентських робіт;
- інше.

Було поставлено завдання створення інформаційної системи (системи підтримки прийняття рішень), яка сприяла б формуванню показників діяльності та проводила перевірку відповідності цих показників галузям освітньої діяльності або науковим спеціальностям.

Розв'язання задач

Формування вхідних даних для розрахунку показників може бути проведено:

- безпосередньо користувачем шляхом завантаження файлів зі списками наукових праць тощо;
- шляхом автоматизованого аналізу інформації з Інтернету.

Другий варіант передбачає використання даних Гугл-Академії [2], де кожен викладач, як правило, має профіль зі списком наукових інтересів у вигляді ключових слів (словосполучень) і списком публікацій, а також автоматично розрахованим числом цитувань, індексом Хірша тощо.

Як система класифікації наукових спеціальностей (фактор об'єктивності) обрано систему Dimensions [3], яка індексує понад 110 мільйонів публікацій і в якій можна за ключовими словами визначити їх відповідність галузі та спеціальності зі списку ANZSRC (Australian and New Zealand Standard Research Classification). У цьому списку варіанта 2008 року наука була розділена на 22 області (Divisions) та 154 спеціальності (Research Groups). Нещодавно система Dimensions перейшла на стандарт ANZSRC-2020, де присутні вже 23 області та 213 спеціальностей.

Приклади моделювання відповідності об'єктів науково-освітньої діяльності науковій категорії можна знайти в [4] – система для підбору наукового керівника, яка використовує ряд джерел інформації, у тому числі – інтереси кандидатів з їх профілів у Google Scholar; [5] – інформаційна технологія синтезу наукового профілю інституту або науково-дослідної лабораторії, яка також використовує інтереси вчених з їх профілів у Google Scholar; [6] – алгоритм тематичного моделювання вчених з наукових спеціальностей на основі їх інтересів у профілях у Google Scholar з використанням переліку наукових спеціальностей із системи класифікації наук ANZSRC. Робота С. Д. Штовби та М. В. Петричко [6] максимально докладно описує застосовувані методи (етапи алгоритму тематичного моделювання, використання індексу Жакара при відсіканні «мінорних» позицій), проте в неї відсутній опис розробленого додатка та його можливостей, наводиться приклад лише роботи з ключовими словами із профілю Google Scholar. Крім того, як показник приймається тільки частота входжень в тематичну колекцію по кожній галузі (спеціальності) без урахування підсумовування по всіх показниках досліджуваного об'єкта.

Особливостями нової створеної системи є наступні:

- завантаження даних як з мережі Інтернет («парсинг» Google-Scholar-профілю наукового співробітника з поділом на список ключових слів і список публікацій), так і з підготовленої таблиці даних;
- робота як з аналізом списку ключових слів (до 100), і з аналізом списку публікацій (до 500);
- введення показників (результатів запиту) із системи Dimensions як в автоматичному режимі, так і вручну;
- розрахунок показників за різними формулами (абсолютні значення числа знайдених входжень; частка кожного значення стосовно суми всіх показників досліджуваного об'єкта; частота входження в тематичну колекцію по кожній галузі або спеціальності; частка кожної частоти входження до суми всіх часток досліджуваного об'єкта);
- аналіз результатів розрахунків як у кожній позиції (парі ключових слів чи наукової публікації) досліджуваного об'єкта, і загального показника (який, своєю чергою, можна розрахувати або як середнє арифметичне, або як суму значень);
- окремий аналіз результатів за галузями, за спеціальностями або за максимальним показником спеціальності в кожній галузі;
- «скорочення» результатів («відсікання хвостів») або за мінімальним рівнем, або за максимальним числом позицій, з можливістю нормалізації підсумку (приведення суми до одиниці);
- перемикання між стандартами ANZSRC-2008 та ANZSRC-2020;

– переклад розподілу за галузями та спеціальностями ANZSRC до таблиці ISCED-F або до стандарту, прийнятого в Україні;

– розрахунок показників по всій групі об'єктів та ранжування об'єктів у цій групі за обраною галуззю чи спеціальністю.

При реалізації системи враховувалося, що Dimensions зберігає та обробляє інформацію англійською мовою, а більшість інформації викладачів – українською та іншими мовами. Тому було передбачено автоматичне переведення інформації.

Таким чином, розроблена система складається з шести модулів.

«Попередній» («Google Scholar») – підключення до сайту Google Scholar та завантаження даних за списком (рис. 1). У разі успішного завантаження управління передається другому модулю. Передбачається, що цей етап функціонуватиме менше (рідше) за інших.



Рисунок 1 – Підключення до сайту Google Scholar та завантаження даних за списком

При успішному завантаженні GS-профілю користувач може виконати сортування даних по іншому полю, розширити діапазон завантажених публікацій тощо, після чого оновити дані для збереження (рис. 2).

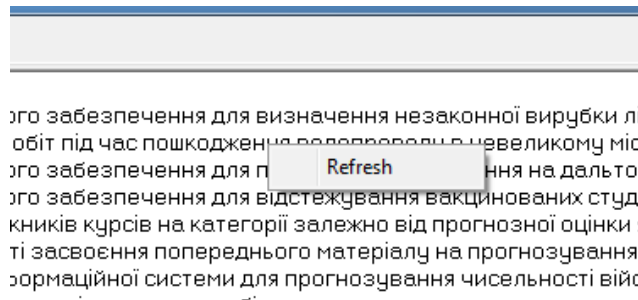


Рисунок 2 – Оновлення даних

У розділі «Робота з даними» («Translating») користувач може побачити завантажені на попередньому етапі дані (статистика вченого, список ключових слів, список статей), відредагувати їх (якщо це необхідно), завантажити нові дані – як у вигляді нового об'єкта, і у вигляді додаткового списку публікацій (рис. 3). Основне завдання цього модуля – переклад англійською мовою [7], отримання англійської копії всіх даних (кнопка Trans1 – переклад ключових слів, кнопка Trans2 – переклад списку публікацій). Отриманий автоматичний переклад може відредагувати користувач.

«Основна робота» («Main Work») – запуск звернення до бази Dimensions та розрахунок за кожним рядком (ключові слова чи назви статей), обробка результату кожного звернення. Завантаження показників із системи Dimensions відбувається один раз («QueryAuto») – циклічне автоматичне запитання даних по кожному рядку; «QueryHand» – «ручне» завантаження, якщо ваша система не дозволить автоматичний доступ).

№	Ім'я	Титулярні	Стр.	Стор.	Стор.	Стор.	Стор.
1	Олександр Рибак	Доклад № 100	87	8	9	2	1 - 100
2	Тетяна Рибак	Доклад № 29	29	3	2	2	1 - 29
3	Володимир Гіт	Доклад № 88	26	4	2	1	1 - 26
4	Наталія Іванівна	Доклад № 22	2	1	1	3	3 - 22
5	Людмила Миколаївна	Доклад № 29	4	3	1	3	3 - 29
6	Катерина Іванівна	Доклад № 6	6	2	2	3	3 - 6
7	Катерина Гіт	Доклад № 11	19	1	1	1	1 - 19
8	Людмила Миколаївна	Доклад № 29	4	3	1	3	3 - 29
9	Тетяна Рибак	Доклад № 29	29	3	2	2	1 - 29

Рисунок 3 – Робота з даними

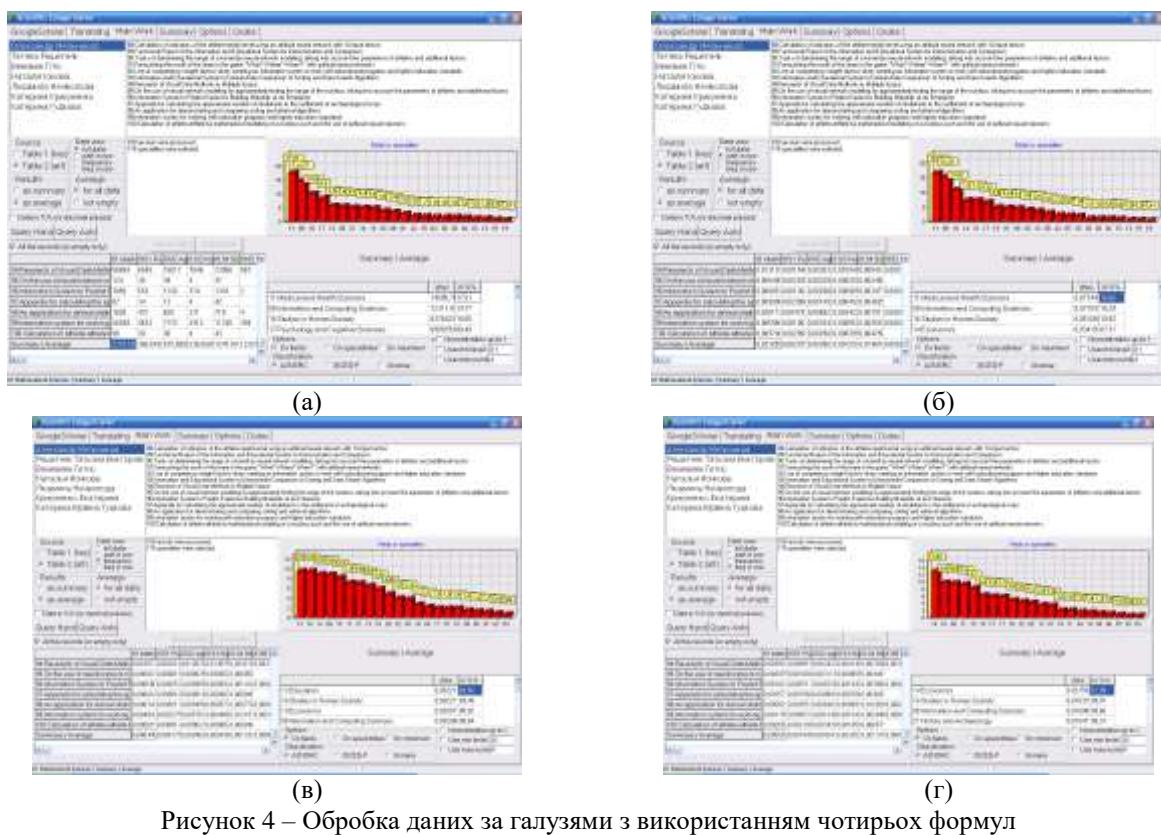


Рисунок 4 – Обробка даних за галузями з використанням чотирьох формул

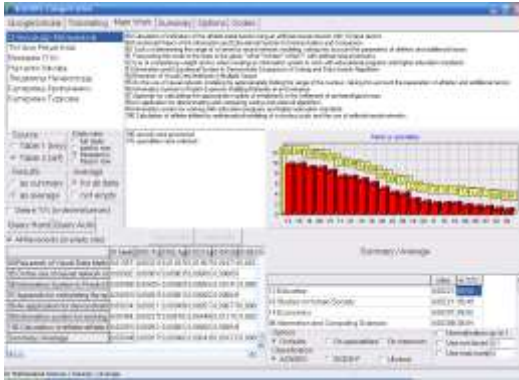
На рис. 4 наведено приклади розрахунку за одним викладачем, але з використанням різних формул розрахунку:

- а) абсолютні значення знайдених входжень, повні дані;
- б) частка кожного значення щодо суми всіх показників рядка (в такому випадку – середнього показника по всіх публікаціях викладача);
- в) частота входжень у тематичну колекцію з кожної галузі;
- г) частка кожної частоти входжень щодо суми всіх часток рядка.

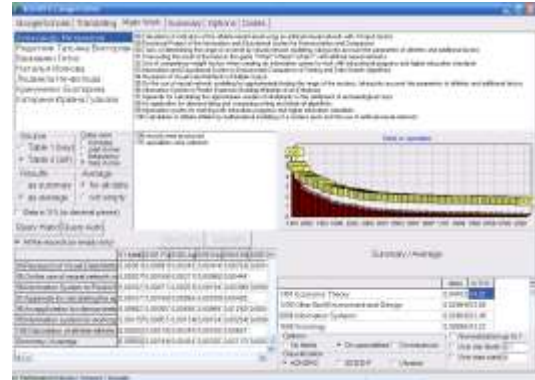
Як видно, ранжування галузей трохи змінюється в залежності від обраної формули розрахунку.

На рис. 5 наведено аналогічні розрахунки за одним викладачем, але вже не за галузями, а за науковими спеціальностями. Для скорочення представленої інформації обмежилися розрахунками за часткою кожного значення щодо суми всіх показників рядка (рис. 5, а) та часткою кожної частоти входжень щодо суми всіх часток рядка (рис. 5, б).

На рис. 6 наведено приклади перекладу на український стандарт галузей (а) та спеціальностей (б), обрано галузь знань, що цікавить нас (12 – Інформаційні технології) та відповідні їй спеціальності.

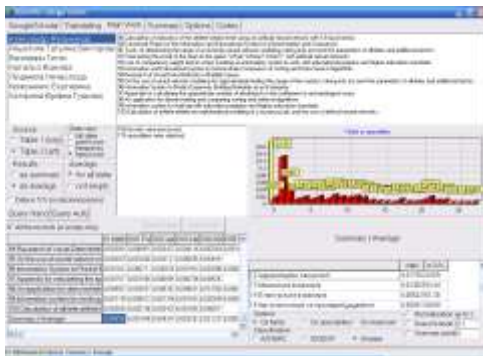


(a)



(б)

Рисунок 5 – Обробка даних за спеціальностями з використанням двох формул

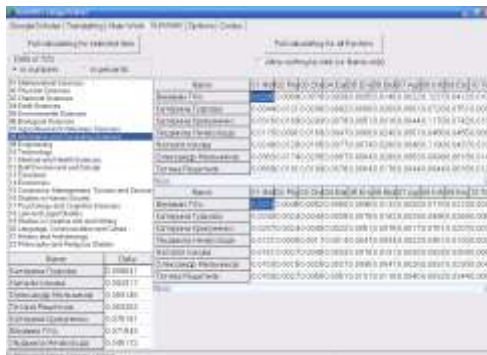


(a)

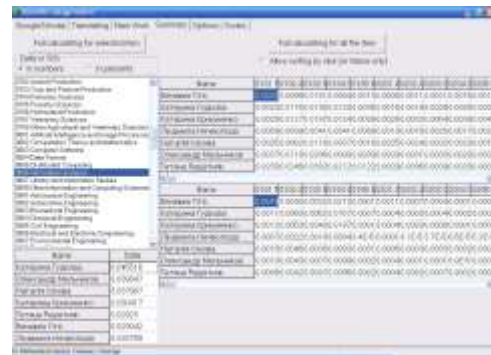


(б)

Рисунок 6 – Обробка даних за галузями (а) та спеціальностями (б) українського стандарту



(a)



(б)

Рисунок 7 – Зведення в таблицю результатів за галузями (а) та спеціальностями (б) стандарту ANZSRC

Розділ «Підбиття підсумків» («Summary») надає можливість зведення розрахунків в єдину таблицю - як за обраною зі списку галузі (спеціальності), так і по всіх галузях (спеціальностям) обраного стандарту.

На рис. 7 наведено приклади відомості в таблицю за стандартом ANZSRC галузей (а) та спеціальностей

Розділ «Коригування параметрів» («Options») дозволяє змінити імена файлів даних, подивитися і при необхідності оновити (кнопки «Get Data» і «Data Update») таблицю кількості записів, що є в Dimensions, по кожній позиції (для розрахунку часток) – рис. 8.

Розділ «Стандарти» («Codes») дозволяє змінити подивитися та перевірити поля доступних у системі стандартів: наразі це ANZSRC, ISCED-F та прийнятий в Україні (рис. 9). Докладніше про роботу цього модуля – у [8-9].

Field ID	Field Name	Value
url	url	
url_name	url_name	
url_page	url_page	
url_pdf	url_pdf	
url_fulltext	url_fulltext	

Field ID	Field Name	Value
01	Mathematical Sciences	1257808
02	Physical Sciences	3810619
03	Applied Mathematics	1216609
04	Numerical and Computational Mathematics	88281
05	Statistics	89136
06	Mathematical Physics	181430
07	Physical Sciences	3888849
08	Astronomical and Space Sciences	386723
09	Astronomy, Astrophysics, Planets and Planets	150371
10	Classical Physics	847234
11	Condensed Matter Physics	84883
12	Optical Physics	82982
13	Quantum Physics	184814
14	Other Physical Sciences	8411623
15	Chemical Sciences	676244
16	Analytical Chemistry	88481
17	Inorganic Chemistry	82282
18	Macromolecular and Materials Chemistry	161766
19	Medical and Biomedical Chemistry	67644
20	Organic Chemistry	171783
21	Physical Chemistry (not Biophysical)	16743

Рисунок 8 – Коригування параметрів та таблиця числа записів у Dimensions

(a)

(б)

Рисунок 9 – Таблиця відповідності (а) та розподілу показників (б) за галузями та спеціальностями різних стандартів

Можливі додаткові опції: знаходження кореляції між інтересами викладача та списком його публікацій (наскільки вони збігаються).

Висновки

Було розглянуто задачу віднесення об'єкта науково-освітньої діяльності (викладача, кафедри, студентської групи тощо) до певної науково-освітньої галузі знань або спеціальності, поставлено та розв'язано задачу створення інформаційної системи (системи підтримки прийняття рішень), яка сприяла б формуванню показників діяльності та проводила перевірку відповідності цих показників галузям освітньої діяльності або науковим спеціальностям. Перелік результатів діяльності об'єкта може при цьому міститися у наукових публікаціях, у ключових словах, якими науковець описує свої наукові інтереси, або в темах кваліфікаційних студентських робіт. Основними інформаційними ресурсами є база категоризованих наукових публікацій із системи Dimensions та інформація про науковців у системі Google Scholar. Використовуються системи класифікації ANZSRC-2008, ANZSRC-2020, ISCED-F та стандарт, прийнятий в Україні.

Описано приклади роботи розробленої системи, розрахунок показників за різними формулами (абсолютні значення числа знайдених входжень; частка кожного значення стосовно суми всіх показників досліджуваного об'єкта; частота входження в тематичну колекцію по кожній галузі або спеціальності; частка кожної частоти входження до суми всіх часток досліджуваного об'єкта); аналіз результатів розрахунків як у кожній позиції (парі ключових слів чи наукової публікації) досліджуваного об'єкта, і загального показника (який, своєю чергою, можна розрахувати або як середнє арифметичне, або як суму значень); «скорочення» результатів («відсікання хвостів») або за мінімальним рівнем, або за максимальним числом позицій, з можливістю нормалізації підсумку (приведення суми до одиниці); переклад розподілу за галузями та спеціальностями ANZSRC до таблиці ISCED-F або до стандарту, прийнятого в Україні; розрахунок показників по всій групі об'єктів та ранжування об'єктів у цій групі за обраною галуззю чи спеціальністю.

Створена система може бути використана при акредитації освітніх програм, при виборі рецензента для наукової статті або студентської конкурсної наукової праці, іноді – при виборі керівника кваліфікаційної роботи тощо.

Список літератури

- [1] Постанова Кабінету міністрів України від 29.04.2015р. № 266. «Про затвердження переліку галузей знань і спеціальностей, за якими здійснюється підготовка здобувачів вищої освіти». [Електронний ресурс]. Режим доступу: <https://zakon.rada.gov.ua/laws/show/266-2015-%D0%BF>. Дата звернення: Бер. 15, 2023.
- [2] Google Академія [Електронний ресурс]. Режим доступу: <https://scholar.google.com.ua/>. Дата звернення: Бер. 15, 2023.
- [3] Dimensions [Електронний ресурс]. Режим доступу: <http://app.dimensions.ai/>. Дата звернення: Бер. 15, 2023.
- [4] Rahdari et al., «Grapevine: A profile-based exploratory search and recommendation system for finding research advisors», *Proceedings of the Association for Information Science and Technology*, 57(1), e271, 2020. doi: 10.1002/pra2.271.
- [5] J. Saad-Falcon, O. Shaikh, Z.J. Wang, A.P. Wright, S. Richardson, and D.H. Chau, «PeopleMap: Visualization Tool for Mapping Out Researchers using Natural Language Processing», arXiv preprint, arXiv:2006.06105, 2020.
- [6] С. Д. Штовба, М. В. Петричко, Тематичне моделювання науковців на основі їх інтересів у Google Scholar, *Системні дослідження та інформаційні технології*, 2021, № 2, с. 113–129.
- [7] Google перекладач [Електронний ресурс]. Режим доступу: <https://translate.google.com/>. Дата звернення: Бер. 15, 2023.
- [8] О. Ю. Мельников, «Додаток для роботи із системами класифікацій галузей знань та спеціальностей», *Інформаційні технології і автоматизація – 2022: матеріали XV Міжнар. наук.-практ. конф., 20–21 жовт. 2022 р.*, Одеса, 2022, с. 115–118.
- [9] О. Ю. Мельников, «Використання систем класифікацій галузей знань та спеціальностей», *Сучасна освіта – доступність, якість, визнання: збірник наукових праць XIV Міжнародної науково-методичної конференції, 9–11 листопада 2022 року, м. Краматорськ-Тернопіль, за заг. ред. д-ра техн. наук., проф. С. В. Ковалевського і Hon. D. Sc., prof. Dasic Predrag*, Краматорськ, ДДМА, 2022, с. 159–162.

Стаття надійшла: 20.05.2023.

References

- [1] Postanova Kabinetu ministriv Ukrainy vid 29.04.2015r. № 266. «Pro zatverdzhennia pereliku haluzei znan i spetsialnostei, za yakymu zdiisnuetsia pidhotovka zdobuvachiv vyshchoi osvity». [Online]. Available: <https://zakon.rada.gov.ua/laws/show/266-2015-%D0%BF>. Accessed on: Mar. 15, 2023. [in Ukrainian]
- [2] Google Scholar [Online]. Available: <https://scholar.google.com.ua/>. Accessed on: Mar. 15, 2023.
- [3] Dimensions [Online]. Available: <http://app.dimensions.ai/>. Accessed on: Mar. 15, 2023.
- [4] Rahdari et al., «Grapevine: A profile-based exploratory search and recommendation system for finding research advisors», *Proceedings of the Association for Information Science and Technology*, 57(1), e271, 2020. doi: 10.1002/pra2.271.
- [5] J. Saad-Falcon, O. Shaikh, Z.J. Wang, A.P. Wright, S. Richardson, and D.H. Chau, «PeopleMap: Visualization Tool for Mapping Out Researchers using Natural Language Processing», arXiv preprint, arXiv:2006.06105, 2020.
- [6] S. D. Shtovba, M. V. Petrychko, Tematychnе modeliuвання naukovtsiv na osnovi yikh interesiv u Google Scholar, *Systemni doslidzhennia ta informatsiini tekhnolohii*, 2021, № 2, s. 113–129. [in Ukrainian]
- [7] Google perekladach [Online]. Available: <https://translate.google.com/>. Accessed on: Mar. 15, 2023.
- [8] O. Yu. Melnykov, «Dodatok dlia roboty iz systemamy klasyfikatsii haluzei znan ta spetsialnostei», *Informatsiini tekhnolohii i avtomatyzatsiia – 2022: materialy XV Mizhnar. nauk.-prakt. CONF., 20–21 zhovt. 2022 r.*, Odesa, 2022, s. 115–118.
- [9] O. Yu. Melnykov, «Vykorystannia system klasyfikatsii haluzei znan ta spetsialnostei», *Suchasna osvita – dostupnist, yakist, vyznannia: zbirnyk naukovykh prats XIV Mizhnarodnoi naukovometodychnoi konferentsii, 9–11 lystopada 2022 roku, m. Kramatorsk-Ternopil, za zah. red. d-ra tekhn. nauk., prof. S. V. Kovalevskoho i Hon. D. Sc., prof. Dasic Predrag*, Kramatorsk, DDMA, 2022, s. 159–162.

Відомості про авторів

Мельников Олександр Юрійович – кандидат технічних наук, доцент, доцент кафедри інтелектуальних систем прийняття рішень Донбаської державної машинобудівної академії.

O. Yu. Melnykov

**DECISION-MAKING SUPPORT SYSTEM FOR
DETERMINING THE FITNESS OF SCIENTIFIC AND
EDUCATIONAL ACTIVITY OBJECTS TO SCIENTIFIC AND
EDUCATIONAL FIELDS AND SPECIALTIES**

Donbas State Engineering Academy, Kramatorsk