

**Кучанський Олександр Юрійович**

Кандидат технічних наук, доцент, доцент кафедри інформаційних систем та технологій,

[orcid.org/0000-0003-1277-8031](https://orcid.org/0000-0003-1277-8031)

Київський національний університет імені Тараса Шевченка, Київ

**ЗАДАЧА ІДЕНТИФІКАЦІЇ ПРЕДМЕТНИХ НАУКОВИХ ПРОСТОРІВ**

**Анотація.** Інформаційний науковий простір незалежно від географічних або фізичних вимірів можна структурувати за різними ознаками. Важливою ознакою, яка часто є в основі коректного аналізу наукових мереж, є спільний напрям наукових досліджень. Розглянуто задачу ідентифікації наукових просторів суб'єктів наукової діяльності, яких об'єднує подібне спрямування наукових праць. Суб'єктами предметних наукових просторів є науковці, які, як правило, об'єднуються у наукові спільноти, що визначаються співавторством спільних наукових публікацій, або індивідуальні науковці, які працюють у визначеному напрямі, проте не цитують та не публікують спільних з іншими науковцями статей. Зв'язки між науковцями, які складають такий науковий простір, задаються через мережу наукової співпраці або мережу цитування. Ідентифікація предметних наукових просторів необхідна для: визначення того, в яких пріоритетних напрямках досліджень працює кожен науковець індивідуально, який його вклад в ці напрями; дослідження ідентифікаторів суб'єктів такого наукового напрямку; оцінювання потенціалу предметного наукового простору, враховуючи динаміку розвитку суб'єктів, які складають цей простір. Для ідентифікації предметних наукових просторів необхідно виконати низку етапів: задати метричний простір та розглянути методи визначення відстані між публікаціями; провести кластерний аналіз публікацій науковців за заданими критеріями та побудувати кластери подібних наукових публікацій; проаналізувати результати кластеризації та об'єднати близькі кластери, провести найменування кластерів, тобто встановити відповідність між кластерами публікацій та відповідними їх назвами; побудувати на основі ідентифікованих кластерів предметні наукові простори, що складатимуться із суб'єктів наукової діяльності; визначити ключові характеристики кожного предметного наукового простору. В роботі описано ці етапи та визначено методи, що можуть бути застосовані для якісної ідентифікації предметних наукових просторів.

**Ключові слова:** науковий простір; кластерний аналіз публікацій; мережа наукової співпраці; мережа цитування; наукометрія

**Вступ**

Незалежно від географічних або фізичних вимірів інформаційний науковий простір можна структурувати за різними ознаками. Важливою ознакою, яка часто є в основі коректного аналізу наукових мереж, є спільний напрям наукових досліджень.

В роботі розглянуто задачу ідентифікації наукових просторів суб'єктів наукової діяльності, яких об'єднує подібне спрямування наукових праць або спільний предметний напрям. Такий простір назвемо предметним науковим простором. Суб'єктами таких просторів є науковці, які як правило, об'єднуються у наукові спільноти, що визначаються співавторством спільних наукових публікацій, або індивідуальні науковці, які працюють у визначеному напрямі, проте не цитують і не публікують спільних з іншими науковцями статей. Зв'язки між науковцями, які складають такий

науковий простір задаються через мережу наукової співпраці або мережу цитування. Ідентифікації предметних наукових просторів необхідна для:

1. Визначення того, в яких пріоритетних напрямках досліджень працює кожен науковець індивідуально, а також який його вклад в ці напрями.
2. Дослідження ідентифікаторів суб'єктів такого наукового напрямку.
3. Оцінювання потенціалу предметного наукового простору, враховуючи динаміку розвитку суб'єктів, які складають цей простір.

Кожний збірник наукових праць має власний перелік предметних напрямів (subject area). Отже, кожен науковець, що публікується в цих збірниках, може отримувати до власного переліку предметних напрямів частину напрямів, за якими визначається збірник. Такий метод ідентифікації напрямів наукової діяльності для науковців має недоліки, які пов'язані з тим, що до переліку предметних напрямів науковця можуть потрапити сторонні напрями, які

стосуються науковця або опосередковано і безпосередньо не відповідають його компетентностям, або взагалі не стосуються його діяльності. Наприклад, збірник наукових праць Eastern-European Journal of Enterprise Technologies (видавець PC Technology Center, ISSN:1729-3774, E-ISSN:1729-4061) має такий перелік предметних напрямів згідно з базою Scopus [1]: промислова та виробнича інженерія; машинобудування; енергетика та енергетичні технології; електротехнічне та електронне машинобудування; прикладна математика; бізнес, управління та облік: управління технологіями та інноваціями; управління та інженерія систем; комп'ютерні науки. Тобто спектр предметних напрямів доволі широкий.

Якщо розглянемо профіль автора Alexander Kuchansky [2], то з 16 опублікованих в базі статей і 102 їх цитувань (станом на 20.05.2020) база Scopus визначила для нього такі предметні напрями: інженерія; комп'ютерні науки; математика; бізнес, управління та облік; енергетика; фізика та астрономія; соціальні науки; наука прийняття рішень. З цього переліку у автора немає жодної статті, яка б навіть опосередковано стосувалася енергетики, фізики та астрономії. Подібна ситуація з формуванням предметних напрямів є у переважній більшості науковців, профілі яких ведуться в цій базі. Отже, можна зробити висновок, що такий метод формування переліку предметних напрямів науковця може бути тільки допоміжним інструментом, проте без безпосереднього контактування з науковцем для визначення його пріоритетних наукових компетентностей він не може братися за основу.

Інший підхід реалізовано на базі відкритого програмного забезпечення Mendeley для керування бібліографічною інформацією, яке дає змогу зберігати та переглядати дослідницькі праці. В профілі автора є можливість переглянути статистику прочитання публікацій, враховуючи інформацію про науковий ступінь або статус особи, яка переглянула публікацію та її пріоритетний предметний напрям. Наприклад, публікація автора Alexander Kuchansky [3] з назвою Learning space conceptual model for computing games developers, яка опублікована за матеріалами конференції «2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2018» має станом на 25.04.2020 – 11 завантажень науковців, які мають профілі у Mendeley та 38 авторів, які профілів не мають. Серед тих, хто має профілі: 60% мають ступінь PhD, 20% професори і 20% наукові співробітники. Також серед тих, хто має профілі у Mendeley: 56% ідентифіковані як спеціалісти у галузі комп'ютерних наук, 22% – в галузі соціальних наук, 11% – економіка, фінанси та економетрика, 11% – інженерія. Таким чином,

знаючи пріоритетні напрями науковців, які вони самі визначили при реєстрації в системі, можна вважати, що якщо статтю здебільшого цікавляться спеціалісти у конкретній галузі (наприклад, комп'ютерних наук), то стаття стосується цієї галузі тією або іншою мірою.

Зрозуміло, що як автоматичне визначення предметних напрямів, так і автоматизоване має недоліки. Неможливо чітко ідентифікувати предметні напрями науковців, не враховуючи розмитість внаслідок міждисциплінарного зв'язку між цими напрямками. Тому актуальність дослідження цього питання є безумовною і потребує як теоретичних, так і практичних обґрунтувань.

## Мета статті

Мета – опис етапів ідентифікації предметних наукових просторів та визначення методів, що можуть бути застосовані для якісної ідентифікації предметних наукових просторів.

## Виклад основного матеріалу

### Ідентифікація предметних наукових просторів

Нехай  $W_s$  – деякий предметний науковий простір,  $W_s \subset W$ , де  $W$  – загальний освітньо-науковий простір. Нехай  $A = \{a_1, a_2, \dots, a_n\}$  – множина всіх суб'єктів наукової діяльності або науковців різних наукових галузей, які публікують свої наукові праці,  $n$  – кількість науковців. Позначимо через  $P = \{p_1, p_2, \dots, p_m\}$  – множину публікацій, які опубліковані цими науковцями,  $m$  – кількість публікацій.

Вважатимемо, що задано декартовий добуток  $A \times P = \{(a, p) \mid a \in A \wedge p \in P\}$  та бінарне відношення  $U \subset A \times P$ , що відображає авторство публікацій  $p \in P$ . Також задамо бінарне відношення  $C \subset P \times P$ , що визначає цитування всіх публікацій  $p \in P$ ,

$$P \times P = \{(p_i, p_j) \mid p_i, p_j \in P, i \neq j\}.$$

Тоді множину всіх публікацій науковця  $a_i$ ,  $i = \overline{1, n}$  позначимо через [4]:

$$P(a_i) = \{p_j \in P \mid (a_i, p_j) \in U\}, \quad i = \overline{1, n}, \quad j = \overline{1, m}.$$

Визначимо множину публікацій, які цитує кожен науковець  $a_i$ ,  $i = \overline{1, n}$ , –  $\overline{C}(a_i)$  та множину публікацій, у яких цитуються публікації науковця –  $C(a_i)$ , таким чином:

$$\overline{C}(a_i) = \{p_j \in P \mid (p_y, p_j) \in C, p_y \in P(a_i), y = \overline{1, m}\}$$

Для кожної публікації  $p_j$  розглянемо множину її авторів:

$$A(p_j) = \{a_i \in A \mid (a_i, p_j) \in U\}, \quad i = \overline{1, n}, \quad j = \overline{1, m},$$

а також множину публікацій, які цитує задана публікація  $p_j - \overline{C}(p_j)$  та множину публікацій, у яких цитується публікація  $p_j - C(p_j)$

$$\overline{C}(p_j) = \{p_y \in P \mid (p_j, p_y) \in C, y = \overline{1, m}\},$$

$$C(p_j) = \{p_y \in P \mid (p_y, p_j) \in C, y = \overline{1, m}\}, \quad j = \overline{1, m}.$$

Задача полягає у побудові таких предметних просторів  $W_s$ , які б склалися з науковців, які працюють у подібному науковому напрямі. Для цього пропонується виконати низку етапів:

1. Задати метричний простір та розглянути методи визначення відстані між публікаціями.

2. Провести кластерний аналіз публікацій науковців за заданими критеріями та побудувати кластери подібних наукових публікацій.

3. Проаналізувати результати кластеризації. За потреби об'єднати близькі кластери.

4. Провести найменування кластерів, тобто встановити відповідність між кластерами публікацій та відповідними їх назвами. Побудувати на основі ідентифікованих кластерів предметних наукових просторів, що складатимуться із суб'єктів наукової діяльності.

5. Визначити ключові характеристики кожного предметного наукового простору та побудувати, якщо це можливо, предметні мережі наукової співпраці за суб'єктами наукової діяльності, що до цього простору входять.

6. Прогнозування динаміки розвитку та потенціалу предметного наукового простору.

Нехай задано метричний простір  $(P, g)$ , який складається з множини публікацій  $P$  та деякої відстані  $g$ , визначеної для будь-якої пари елементів цієї множини. Відстань  $g$  визначається як відображення з множини декартового квадрату множини  $P$  на множину дійсних чисел, тобто:

$$g: P \times P \rightarrow \mathbf{R},$$

де  $\mathbf{R}$  – множина дійсних чисел.

Відстань між публікаціями  $g$  є невід'ємною, дійсною функцією  $g(p_i, p_j) \geq 0$ ,  $i = \overline{1, m}$ ,  $j = \overline{1, m}$ , яка визначається для  $\forall p_i, p_j \in P$ . Для функції  $g(p_i, p_j)$ ,  $i \neq j$  виконуються аксіоми тотожності, симетрії та нерівності трикутника [4; 5].

#### Задання метричного простору та методи визначення відстані між публікаціями науковців

*Розрахунок відстані між публікаціями на основі довжини найкоротшого шляху між вершинами графа  $(P, C)$ , які відображають ці публікації.* Відстань  $g(p_i, p_j)$  між довільними вершинами  $p_i, p_j \in P$  визначатиметься як довжина мінімального

шляху з вершини  $p_i$  у вершину  $p_j$ , якщо він існує.

Вважатимемо, що ваги дуг рівні 1, тобто вагова функція графа  $(P, C)$   $f: C \rightarrow \{1\}$  і довжина шляху з вершини  $p_i$  у вершину  $p_j$  визначатиметься так:

$$r(p_i, p_j) = \sum_{i=0}^{k-1} f(c_{y, y+1}).$$

Якщо між вершинами  $p_i$  і  $p_j$  шляху немає, то довжину визначимо так:  $r(p_i, p_j) = m + 1$ .

*Розрахунок відстані між публікаціями на основі ступеня близькості за змістом анотацій цих публікацій.* Для визначення ступеня близькості публікацій за змістом будемо здійснювати порівняння не всього тексту публікацій, а тільки анотацій цих публікацій. Такий метод засновано на припущенні, що анотації наукових публікацій, які належать до одного і того ж наукового напрямку будуть достатньо близькі за змістом та містити одні і ті ж поняття або ключові слова [6].

#### Кластерний аналіз публікацій науковців.

Нехай задано метричний простір  $(P, g)$ . Необхідно розбити множину публікацій  $P$  на деяку кількість підмножин, які між собою не перетинаються. Такі підмножини множини  $P$  називаються кластерами. Позначимо множину кластерів через  $Y = \{y_1, y_2, \dots, y_z\}$ , де  $z$  – кількість кластерів, на які розбивається множина  $P$ . Для кластеризації графа використовуємо метод Louvain. Детально метод кластеризації графа описано в роботі [7].

В результаті проведення процедури кластеризації наукових публікацій отримуємо множину кластерів  $Y$ . Потужність множини  $Y$  може бути достатньо великою, що ускладнює подальший аналіз. Одним зі способів вирішення цієї проблеми є укрупнення побудованих кластерів шляхом злиття близьких між собою кластерів з невеликою кількістю елементів [8; 9].

#### Найменування кластерів та побудова предметних наукових просторів.

Після проведення процедури кластеризації графа  $(P, C)$  та об'єднання близьких між собою кластерів, необхідно встановити відповідність між конкретним кластером та вербальною назвою напрямку досліджень, який цей кластер представляє. Тобто, якщо  $Y = \{y_1, y_2, \dots, y_z\}$  – побудована множина кластерів після застосування одного з алгоритмів кластеризації графа  $(P, C)$ , а  $\overline{Y} = \{y_{k_1}, y_{k_2}, \dots, y_{k_v}\}$  – кінцева множина кластерів, яка побудована в результаті виконання алгоритму об'єднання близьких кластерів,  $k_j \in \{1, 2, \dots, z, z+1, \dots, z+v\}$  – індекси елементів кінцевої множини кластерів,

$h = \overline{1, \psi}$ ,  $v$  – кількість об'єднань кластерів при виконанні алгоритму об'єднання,  $\psi$  – кількість елементів кінцевої множини кластерів.

Кожному кластеру  $u_{k_1}, u_{k_2}, \dots, u_{k_\psi}$  покладемо у відповідність певний напрям наукових досліджень. Тобто розглянемо відображення  $\Phi: \overline{Y} \rightarrow V$ , де  $V$  – множина вербальних назв науково-дослідних напрямів. Наприклад, елементами множини  $V$  можуть бути напрями: «Математична фізика», «Теорія оптимізації», «Комп'ютерні науки» тощо. Для встановлення відповідності  $\Phi$  можна використати експертний підхід. У такому разі експерти прийматимуть рішення про встановлення відповідності кожному кластеру відповідного напрямку, базуючись на списку публікацій кластеру та деяких додаткових відомостей, зокрема ключових словах, найбільш вживаних поняттях тощо.

Позначимо через  $V = \{\eta_1, \eta_2, \dots, \eta_\psi\}$  множину напрямів наукових досліджень,  $\psi$  – кількість напрямів наукових досліджень. Як вже зазначалося, ідентифікація напрямів досліджень науковців – це процес встановлення відповідності між конкретним науковцем та науковими напрямами, в яких цей науковець працює та публікує в рамках цих напрямів наукові публікації. Тобто необхідно знайти відображення  $\Lambda: A \rightarrow V$ . Для проведення ідентифікації напрямів досліджень науковців одним зі способів є використання інформації щодо публікаційної активності науковців, враховуючи побудовану множину кластерів наукових напрямів, до яких ці публікації належать. Зрозуміло, що наукові публікації в переважній більшості публікуються зі співавторами.

Особливістю розв'язку задачі є те, що предметні наукові простори перетинаються внаслідок того, що деякі з науковців потрапляють одразу до кількох просторів, тобто їх публікації можуть належати до кількох наукових напрямів. В цьому випадку визначити чітко належність науковця до відповідного предметного наукового простору неможливо і ця належність буде нечіткою.

Позначимо через  $\overline{P}_{k_z}(a_i)$  множину публікацій автора  $a_i$ ,  $i = \overline{1, n}$ , що включена до кластеру  $u_{k_z}$ ,  $z = \overline{1, \psi}$ :

$$\overline{P}_{k_z}(a_i) = \{p_j \in P \mid (a_i, p_j) \in U, p_j \in u_{k_z}\}.$$

Тоді кількість публікацій науковця  $a_i$ ,  $i = \overline{1, n}$  в кожному з кластерів  $u_{k_z}$ ,  $z = \overline{1, \psi}$  визначатимуться:

$$R_{k_z}(a_i) = \text{card}(\overline{P}_{k_z}(a_i)),$$

де  $R_{k_z}(a_i)$  – кількість публікацій науковця  $a_i$  в кластері  $u_{k_z}$ .

Якщо всі публікації належать своїм кластерам, то  $m = \sum_{i=1}^n \sum_{z=1}^{\psi} R_{k_z}(a_i)$  і можна задати множину впорядкованих пар, що складаються з елементів  $a \in A$  і відповідної функції належності  $\mu_{y_z}(a)$ ,  $z = \overline{1, \psi}$ , тобто множини  $E^z = \{(a, \mu_{y_z}(a)) \mid a \in A\}$ , де функція належності  $\mu_{y_z}(a)$  вказує на те, в якій мірі елемент  $a \in A$ , тобто науковець, належить нечіткій множині  $E^z$ , що відповідає визначеному напрямку наукових досліджень або предметному науковому простору. Функція  $\mu_{y_z}(a)$  приймає значення на деякій множині належностей  $[0, 1]$ .

Носієм нечіткої множини  $E^z$ ,  $\text{sup p } E^z$  називається така множина:

$$\text{sup p } E^z = \{a \mid a \in A, \mu_{y_z}(a) > 0\}.$$

Визначимо функцію  $\mu_{y_z}(a)$  так:

$$\mu_{y_z}(a_i) = \begin{cases} 0, & \text{sup p } E^z \cap a_i = \emptyset, \\ \frac{R_{k_z}(a_i)}{R(a_i)}, & \text{sup p } E^z \cap a_i \neq \emptyset, \end{cases}$$

де  $R(a_i) = \sum_{z=1}^{\psi} R_{k_z}(a_i)$ ,  $i = \overline{1, n}$ .

Значення  $\mu_{y_z}(a_i) = 1$ , коли  $\bigcup_{\substack{z=1, \psi, \\ z \neq i}} \text{sup p } E^z \cap a_i = \emptyset$ ,

причому  $\sum_{z=1}^{\psi} \mu_{y_z}(a_i) = 1$  для фіксованого  $i = \overline{1, n}$ .

Множина напрямів наукових досліджень науковця  $a_i$  визначається так:

$$\Theta(a_i) = \{\Phi(y_{k_h}) = \eta_h, h = \overline{1, \psi} \mid y_{k_h} \cap P(a_i) \neq \emptyset\},$$

де  $\Theta(a_i)$  – множина напрямів наукових досліджень науковця  $a_i$ ,  $i = \overline{1, n}$ ,  $\Phi: \overline{Y} \rightarrow V$ .

Сукупність множин  $\Theta(a_i)$  для всіх науковців є розв'язком задачі ідентифікації наукових напрямів. Отримані результати дають змогу розв'язати споріднену задачу знаходження науковців, що здійснюють дослідження у заданому напрямі.

Для знаходження  $\Phi: \overline{Y} \rightarrow V$  можна використати нечітке експертне оцінювання.

Нехай  $V = \{\eta_1, \eta_2, \dots, \eta_\psi\}$  – являють собою задалегідь сформований, на основі ключових слів публікацій, перелік термів. Необхідно оцінити змінну «відповідність кластеру до напрямку наукових досліджень». Проводиться експертне опитування і результати записуються у таблицю.



Таблиця – Таблиця для проведення експертного опитування

	ТЕРМИ			...	
E <sub>1</sub>	η <sub>1</sub>				
	η <sub>2</sub>				
	...				
	η <sub>ψ</sub>				
E <sub>2</sub>	η <sub>1</sub>				
	η <sub>2</sub>				
	...				
	η <sub>ψ</sub>				
...	...	...	...	...	...
E <sub>L</sub>	η <sub>1</sub>				
	η <sub>2</sub>				
	...				
	η <sub>ψ</sub>				

У таблиці  $\lambda_{i,j}^k \in \{0,1\}$  – оцінка k-го експерта наявності у елемента і властивості j,  $i, j = \overline{1, \psi}$ . Значення 1 вказує на наявність властивості, 0 – на відсутність. Кластеру  $u_k$  ставимо у відповідність такий напрям наукової діяльності  $\eta_j$ , для якого виконується умова  $\max_j \sum_{k=1}^L \lambda_{i,j}^k$ . Вважаємо, що

кожному кластеру можна поставити у відповідність один напрям. Якщо при розрахунку визначається одразу кілька кластерів, які можна віднести до одного напрямку, то ці кластери об'єднують.

### Висновки

В роботі розглянуто актуальну задачу ідентифікації предметних наукових просторів на основі даних про цитування публікацій суб'єктів наукової діяльності. Побудовано метод, на основі якого можна визначити межі предметних наукових просторів. Тобто автоматично сформувані перелік пріоритетних напрямів наукової діяльності, якими займаються суб'єкти наукової діяльності.

Ідентифіковані предметні наукові простори складають основу для подальшого їх дослідження та оцінювання науковців, які до цих предметних просторів належать. Результати ідентифікації можуть бути використані для розв'язання задачі вибору партнерів для наукової співпраці, створення спільних проєктних груп, консорціумів університетів тощо [10]. Також отримані результати можуть бути використані для забезпечення навчання та проблеми впровадження міжнародних стандартів оцінювання компетенцій керівників наукових проєктів та програм [11 – 14].

### Список літератури

1. Scopus. Eastern-European Journal of Enterprise Technologies. [Електронний ресурс], URL: <https://www.scopus.com/sourceid/21100450083>
2. Scopus. Kuchansky, Alexander [Електронний ресурс], URL: <https://www.scopus.com/authid/detail.uri?authorId=57190488151>
3. Mendeley. Kuchansky, Alexander [Електронний ресурс], URL: <https://www.mendeley.com/authors/57190488151/>
4. Biloshchytskyi A., Kuchansky A., Andrashko Yu., Biloshchytska S., Kuzka O., Terentyev O., (2017). Evaluation methods of the results of scientific research activity of scientists based on the analysis of publication citations. Eastern-European Journal of Enterprise Technologies, 3(2 (87)), 4 – 10. doi: 10.15587/1729-4061.2017.103651
5. Kuchansky A., Andrashko Yu., Biloshchytskyi A., Danchenko O., Ilarionov O., Vatskel I., Honcharenko T. The method for evaluation of educational environment subjects' performance based on the calculation of volumes of M-simplexes. Eastern-European Journal of Enterprise Technologies. 2018. 2(4(92)). P. 15 – 25. doi: 10.15587/1729-4061.2018.126287
6. Lizunov P., Biloshchytskyi A., Kuchansky A., Andrashko Yu., Biloshchytska S. Improvement of the method for scientific publications clustering based on n-gram analysis and fuzzy method for selecting research partners. Eastern-European Journal of Enterprise Technologies. 4(4(100)). 2019. P. 6–14. doi: <https://doi.org/10.15587/1729-4061.2019.175139>
7. Blondel V., Guillaume J., Lambriotte R., Lefebvre E. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. 2008, P10008. doi:10.1088/1742-5468/2008/10/P10008
8. Biloshchytskyi A., Kuchansky A., Andrashko Yu., Biloshchytska S., Kuzka O., Shabala Ye., Lyashchenko T. A method for the identification of scientists' research areas based on a cluster analysis of scientific publications // Eastern-European Journal of Enterprise Technologies. – 2017. – Vol. 5. – Issue 2 (89). – P. 4 – 10. doi:10.15587/1729-4061.2017.112323
9. Biloshchytskyi A., Kuchansky A., Andrashko Yu., Paliy S., Biloshchytska S., Bronin S., et al. Development of technical component of the methodology for project-vector management of educational environment. Eastern-European Journal of Enterprise Technologies. 2018. Vol. 2/2 (92), 4 – 13, doi: 10.15587/1729-4061.2018.126301
10. Xu, H., Kuchansky, A., (2019). The problem of choice of partners for organization of cooperation in the framework of scientist of scientific and educational projects. Scientific Bulletin of Uzhhorod National University. Series of Mathematics and Informatics, 2(35), 110 – 119.
11. Morozov V., Stechenko G., Kolomiets A. "Learning Through Practice" in IT Management Projects Master Program Implementation Approach // 2017 9<sup>th</sup> IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS): Bucharest. 2017. 2. P. 935 – 939.

12. Morozov V., Kalnichenko O., Liubyma I. Projects change management in based on the projects configuration management for developing complex projects // 2017 9<sup>th</sup> IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS): Bucharest. 2017. 2. P. 939 – 942.

13. Kolesnikov O., Gogunskii V., Kolesnikova K., Lukianov D., Olekh T. Development of the model of interaction among the project, team of project and project environment in project system // Eastern-European Journal of Enterprise Technologies. 2016. 5(9(83)), P. 20–26. DOI: 10.15587/1729-4061.2016.80769

14. Kolesnikova K., Lukianov D., Gogunskii V., Iakovenko V., Oborska G., Negri A., Koliada A., Dmitrenko K., Olekh T., Bespanskaya-Paulenka K. (2017). Communication management in social networks for the actualization of publications in the world scientific community on the example of the network researchgate // Eastern-European Journal Of Enterprise Technologies, 4(3 (88)), 27 – 35, doi:10.15587/1729-4061.2017.108589

Стаття надійшла до редколегії 28.04.2020

### Kuchansky Alexander

PhD (Eng.), Associate Professor, Department of Information Systems and Technologies, orcid.org/0000-0003-1277-8031  
Taras Shevchenko National University of Kyiv, Kyiv

## THE PROBLEM OF SCIENTIFIC SUBJECT SPACES IDENTIFICATION

**Abstract.** Information science space, regardless of geographical or physical dimensions, can be structured on various grounds. An important feature that is the basis of a correct analysis of scientific networks is the common direction of scientific research. The paper considers the problem of identification of scientific spaces of subjects of scientific activity, which are united by a similar direction of scientific works. Subjects of scientific subject spaces are scientists, who usually unite in scientific communities, which are determined by the co-authorship of joint scientific publications, or individual scientists who work in a certain direction, but do not cite or publish joint articles with other scientists. The links between the scientists that make up such a scientific space are established through a network of scientific collaboration or a network of citations. Identification of scientific subject spaces is necessary to: determine in which priority areas of research each scientist works individually and what is his contribution to these areas; research of identifiers of subjects of such scientific direction; assessment of the potential of the scientific subject space, taking into account the dynamics of development of the subjects that make up this space. To identification scientific subject spaces, it is necessary to perform a number of stages: set the metric space and consider methods for establishing the distance between publications; to conduct a cluster analysis of publications of scientists according to the given criteria and to build clusters of similar scientific publications; analyze the results of clustering and combine close clusters, name clusters, i.e., establish a correspondence between clusters of publications and their corresponding names; to build on the basis of identified clusters of scientific subject spaces, which will consist of subjects of scientific activity; identify the key characteristics of each scientific subject space. The article describes these stages and identifies methods that can be used for qualitative identification of scientific subject spaces.

**Keywords:** scientific space; cluster analysis of publications; scientific collaboration network; citation network; scientometrics

### References

1. Scopus. (2020). Eastern-European Journal of Enterprise Technologies. <https://www.scopus.com/sourceid/21100450083> Accessed 20 Apr 2020.
2. Scopus. (2020). Kuchansky, Alexander. <https://www.scopus.com/authid/detail.uri?authorId=57190488151> Accessed 20 Apr 2020.
3. Mendeley. (2020). Kuchansky, Alexander. URL: <https://www.mendeley.com/authors/57190488151/>
4. Biloshchytskyi, A., Kuchansky, A., Andrashko, Yu., Biloshchytska, S., Kuzka, O. & Terentyev, O. (2017). Evaluation methods of the results of scientific research activity of scientists based on the analysis of publication citations. Eastern-European Journal of Enterprise Technologies, 3(2 (87)), 4–10. doi: 10.15587/1729-4061.2017.103651
5. Kuchansky, A., Andrashko, Yu., Biloshchytskyi, A., Danchenko, O., Ilarionov, O., Vatskel, I., & Honcharenko, T. (2018). The method for evaluation of educational environment subjects' performance based on the calculation of volumes of M-simplexes. Eastern-European Journal of Enterprise Technologies, 2(4(92)), 15 – 25. doi: 10.15587/1729-4061.2018.126287
6. Lizunov, P., Biloshchytskyi, A., Kuchansky, A., Andrashko, Yu., & Biloshchytska, S. (2019). Improvement of the method for scientific publications clustering based on n-gram analysis and fuzzy method for selecting research partners. Eastern-European Journal of Enterprise Technologies. 4(4(100)). 6–14. doi: <https://doi.org/10.15587/1729-4061.2019.175139>
7. Blondel, V., Guillaume, J., Lambriotte, R., & Lefebvre, E. (2008) Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008:P10008. doi:10.1088/1742-5468/2008/10/P10008.

8. Biloshchytskyi, A., Kuchansky, A., Andrashko, Yu., Biloshchytska, S., Kuzka, O., Shabala, Ye., & Lyashchenko, T. (2017). A method for the identification of scientists' research areas based on a cluster analysis of scientific publications. *Eastern-European Journal of Enterprise Technologies*, 5(2(89)), 4-10. doi:10.15587/1729-4061.2017.112323
9. Biloshchytskyi, A., Kuchansky, A., Andrashko, Yu., Paliy, S., Biloshchytska, S., Bronin, S., et al. (2018). Development of technical component of the methodology for project-vector management of educational environment. *Eastern-European Journal of Enterprise Technologies*, 2/2 (92), 4–13. doi: 10.15587/1729-4061.2018.126301
10. Xu, H., & Kuchansky, A. (2019). The problem of choice of partners for organization of cooperation in the framework of scientist of scientific and educational projects. *Scientific Bulletin of Uzhhorod National University. Series of Mathematics and Informatics*, 2(35), 110-119.
11. Morozov, V., Stechenko, G., & Kolomiets, A. (2017). "Learning Through Practice" in IT Management Projects Master Program Implementation Approach. 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS): Bucharest, 2, 935–939.
12. Morozov, V., Kalnichenko, O., & Liubyma, I. (2017). Projects change management in based on the projects configuration management for developing complex projects. 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS): Bucharest, 2, 939–942.
13. Kolesnikov, O., Gogunskii, V., Kolesnikova, K., Lukianov, D., & Olekh, T. (2016). Development of the model of interaction among the project, team of project and project environment in project system. *Eastern-European Journal of Enterprise Technologies*, 5(9(83)), 20-26. DOI: 10.15587/1729-4061.2016.80769
14. Kolesnikova, K., Lukianov, D., Gogunskii, V., Iakovenko, V., Oborska, G., Negri, A., Koliada, A., Dmitrenko, K., Olekh, T., & Bepanskaya-Paulenka, K. (2017). Communication management in social networks for the actualization of publications in the world scientific community on the example of the network researchgate. *Eastern-European Journal Of Enterprise Technologies*, 4(3 (88)), 27-35. doi:10.15587/1729-4061.2017.108589

---

#### Посилання на публікацію

- APA Kuchansky, A., (2020). The problem of scientific subject spaces identification. *Management of Development of Complex Systems*, 42, 146 – 152, dx.doi.org/10.32347/2412-9933.2020.42.146-152.
- ДСТУ Кучанський О.Ю. Задача ідентифікації предметних наукових просторів [Текст] / О.Ю. Кучанський // *Управління розвитком складних систем*. – № 42. – 2020. – С. 146 – 152, dx.doi.org/10.32347/2412-9933.2020.42.146-152.