DOI: 10.32347/2412-9933.2025.64.194-200
УДК 004.9

Доманецька Ірина Миколаївна
Кандидат технічних наук, доцент, доцент кафедри інтелектуальних технологій,
https://orcid.org/0000-0002-8629-9933
Київський національний університет імені Тараса Шевченка, Київ
Хроленко Ярослав Олексійович
Аспірант відділу інтелектуальних технологій підтримки прийняття рішень,
https://orcid.org/0009-0004-0641-827X
Інститут проблем реєстрації інформації НАН України, Київ

ОСОБЛИВОСТІ ВИЗНАЧЕННЯ ОПТИМАЛЬНОГО СКЛАДУ ЕКСПЕРТНОЇ ГРУПИ НА ОСНОВІ СЕМАНТИКО-СТАТИСТИЧНОГО ПІДХОДУ

Анотація. *Формування складу експертних груп для оцінювання наукових проєктів, конкурсних заявок та рецензування є критично важливою задачею, що безпосередньо впливає на якість, об'єктивність і доброчесність наукової експертизи. Традиційні підходи до добору експертів, зокрема ранжування за індексом Гірша або класифікація за кодами ASJC/MeSH, мають низку обмежень: вони не враховують семантичну релевантність профілю експерта до тематики конкурсу, ігнорують міждисциплінарні зв'язки та не забезпечують поелементного аналізу тематичного покриття. У відповідь на ці виклики в роботі запропоновано семантико-статистичний підхід до оцінки компетентності експертів, який базується на побудові графів співвживання термінів для кожного концепту конкурсної тематики з використанням відкритого словника концептів OpenAlex. Модель дозволяє здійснювати поелементну оцінку тематичної компетентності експертів на основі їхнього публікаційного профілю, враховуючи як прямі, так і опосередковані семантичні зв'язки. У межах дослідження формалізовано алгоритм побудови локальних графів семантичної близькості для кожного концепту, визначено метрику оцінки компетентності, яка враховує частоту згадування термінів у публікаціях та їхню семантичну віддаленість від ядра теми. На основі отриманих оцінок сформовано матрицю компетентностей, що дозволяє застосовувати методи групової оптимізації для формування збалансованих експертних складів. Розглянуто різні стратегії оптимізації – жадібні алгоритми, цілочисельне програмування, еволюційні методи – та обґрунтовано їх застосовність залежно від розміру задачі та вимог до точності. Запропонований підхід забезпечує прозорість, тематичну чутливість і гнучкість формування експертних груп, дозволяє уникнути дублювання компетенцій та недооцінки вузькоспеціалізованих дослідників. Модель є масштабованою, відтворюваною та придатною для інтеграції в системи автоматизованого добору експертів у наукових конкурсах з перспективою розширення на механізми виявлення конфліктів інтересів та оцінки академічної доброчесності.*

Ключові слова: тематична компетентність; формування експертної групи; семантико-статистичний підхід; OpenAlex Concepts; тематичне покриття; співвживання термінів

Постановка проблеми

Формування складу експертних груп для наукових проєктів, конкурсів, рецензування рукописів чи оцінки грантових заявок є важливою задачею, яка впливає на якість і об'єктивність оцінювання. У контексті міждисциплінарних і прикладних напрямів, вибір експертів ускладнюється необхідністю врахування семантичної близькості профілю експерта до специфіки об'єкту експертизи, забезпечення повного покриття тематики експертизи, збалансованості компетенцій і уникнення суб'єктивності. Традиційні методи, як ранжування за індексом Гірша (h-index) чи класифікація за кодами ASJC (Scopus) або MeSH (PubMed), мають суттєві обмеження: вони не враховують семантичну близькість профілю експерта до тематики конкурсу, ігнорують міждисциплінарні зв'язки та часто призводять до дублювання компетенцій у групі.

Мета статті

Мета статті – проаналізувати особливості семантико-статистичного підходу до формування експертних груп. У центрі уваги – перехід до колективної оптимізації, що враховує семантичну релевантність, тематичну віддаленість та комбіновану оцінку компетентності, що підвищує об'єктивність і збалансованість експертизи.

Виклад основного матеріалу

У сучасних умовах в Україні реалізується широкий спектр дослідницьких, науково-технічних та інноваційних проєктів, результати яких потребують фахової оцінки для прийняття обґрунтованих рішень щодо їхньої наукової новизни, практичної цінності та потенціалу впровадження. Експертне оцінювання виступає ключовим інструментом у цьому процесі. У цьому контексті експертиза постає як інструмент науково обґрунтованого вибору, що поєднує аналітичну оцінку з процедурою прийняття рішень. З огляду на зростання міждисциплінарності досліджень та складність об'єктів оцінювання, особливої актуальності набуває формування збалансованих експертних груп, здатних забезпечити комплексне, релевантне та неупереджене оцінювання проєктних ініціатив.

У сучасній практиці формування експертних груп для оцінювання конкурсних матеріалів або наукових проєктів широко застосовуються підходи, що базуються на оцінках компетентності кандидатів. Найпоширенішими є багатокритеріальні моделі, які дозволяють враховувати різні аспекти профілю експерту [1]. Недоліки традиційних багатокритеріальних підходів до формування експертних груп полягають у низькій тематичній чутливості, втраті структурної інформації, обмеженій інтерпретованості та схильності до формальних викривлень. Такі моделі, як правило, базуються на агрегованих показниках (кількість публікацій, індекс Гірша, досвід), які не відображають змістовну релевантність наукового доробку до конкретної тематики конкурсу. Вони не враховують семантичні зв'язки між поняттями, що обмежує здатність моделі виявляти опосередковану компетентність. Крім того, результати оцінювання складно пояснити, оскільки відсутня деталізація внеску окремих термінів. Це створює ризик переоцінки формально активних, але тематично нерелевантних кандидатів, і водночас – недооцінки молодих дослідників із вузькою, але точною спеціалізацією. У міждисциплінарних конкурсах такі моделі виявляються недостатньо гнучкими, оскільки не дозволяють поелементно аналізувати покриття тематики [2; 3].

Таким чином, традиційні бібліометричні показники, як-от індекс Гірша чи кількість цитувань, мають низку обмежень. Вони не враховують змістовну релевантність публікацій до конкретної тематики, не відображають міждисциплінарну або прикладну експертизу, а також схильні до темпоральних викривлень, що ускладнює об'єктивну оцінку молодих або вузькоспеціалізованих дослідників [4].

У відповідь на ці виклики зростає інтерес до моделей, що інтегрують семантичний аналіз наукового доробку. Такі підходи дозволяють оцінювати тематичну близькість між профілем експерта та ключовими словами конкурсу, використовуючи методи тематичного моделювання, аналізу співвживання термінів, побудови онтологій та семантичних графів [5-8]. Наприклад, у роботі [9] запропоновано метод виявлення релевантних експертних груп на основі семантичної структури термінів і тематичного узгодження між експертами. У поєднанні з методами колективної оптимізації (наприклад, жадібними алгоритмами або моделями максимального покриття) це відкриває можливості для формування збалансованих експертних груп, які забезпечують повне та релевантне охоплення тематики.

На цьому тлі актуальним постає завдання розробки підходу до формування складу експертної групи, що поєднує семантико-статистичну оцінку компетентності з механізмами колективної оптимізації, орієнтованими на максимальне тематичне покриття та міждисциплінарну збалансованість.

У межах запропонованого підходу задача формування складу експертної групи формалізується як задача групової оптимізації, що базується на персоналізованих оцінках компетентності експертів щодо кожного ключового слова, яке описує предмет експертизи.

Узагальнений підхід полягає в тому, щоб:

1) Побудувати для кожного ключового слова з опису предмета експертизи семантико-статистичні графи співвживання термінів на основі даних глобальних науково-інформаційних агрегаторів (наприклад, OpenAlex, Scopus, Web of Science), які вже реалізують структури знань, побудовані за принципами семантичної близькості. В якості словника концептів для опису предмету експертизи та наукових робіт експертів в даній роботі використовується словник концептів OpenAlex [10].

2) Використати ці графи для обчислення індивідуальних оцінок тематичної компетентності експертів щодо кожного окремого терміну (ключового слова, концепту, змістової одиниці, топіку відповідного агрегатора), що входить до опису предмету експертизи чи конкурсної тематики.

3) На основі отриманих оцінок сформувати матрицю компетентностей кандидатів, яка дозволяє здійснити вибір експертів з комплементарними профілями для забезпечення повного та збалансованого покриття тематики конкурсу.

Використання класифікаторів концептів із наукометричних агрегаторів обумовлено тим, що вони вже апробовані в глобальних базах даних, мають ієрархічну структуру та охоплюють

міждисциплінарні зв'язки, що дозволяє точно формалізувати предмет експертизи. Їх використання спрощує побудову графів співвживання термінів, забезпечує відтворюваність оцінки компетентності та дозволяє адаптувати модель до різних типів конкурсів без потреби в ручній класифікації.

Розглянемо особливості формування семантико-статистичнного графу співвживання термінів на основі даних науково-інформаційних агрегаторів. У межах семантико-статистичного підходу до оцінювання компетентності експертів пропонується побудова окремого графа семантично-частотної близькості для кожного концепту, що входить до опису предмету експертизи.

Такий граф моделює локальне семантичне оточення терміну на основі статистики співвживання концептів у глобальному корпусі наукових публікацій, що забезпечує тематичну деталізацію оцінки.

Нехай:

$K = \{k_1, k_2, \dots, k_n\}$ – множина концептів, що описують предмет експертизи, де n – їх загальна кількість.

$E = \{e_1, e_2, \dots, e_m\}$ – множина експертів-претендентів, де m – їх загальна кількість.

Для кожного $k_i \in K$ будується окремий граф

$$G_i = (V_i, Q_i, R_i),$$

де V_i – множина концептів (вершин), що семантично пов'язані з k_i, має багатошарову структуру утворену за рахунок врахування опосередкуваного співвживання концептів; $Q_i \subseteq V_i \times V_i$ – множина умовних ребер між парами співвживаних концептів; $R_i(u, v) \in \mathbb{N}$ – вага ребра, що дорівнює кількості публікацій, у яких концепти u та v зустрічаються разом.

Вхідні дані алгоритму

– $\mathcal{W}$– глобальний корпус наукових публікацій, доступний через агрегатор (наприклад, OpenAlex);

– K – множина концептів, що описують предмет експертизи;

– $\theta \in \mathbb{N}$ – порогове значення частоти співвживання для включення ребра.

– $D \in \mathbb{N}$ – максимальна глибина графа (кількість рівнів розширення), зазвичай $D = 5$.

Алгоритм побудови семантико-статистичного графа G_i для концепту k_i містить кроки:

Крок 1: Ініціалізація

1. Обирається концепт k_i з опису тематики експертизи.
2. Встановлюється початковий рівень $d := 0$.
3. Множина концептів на нульовому рівні у графі G_i містить лише концепт $L_0^{(i)} := \{k_i\}$, тобто граф починається з однієї вершини – ядра.

$$V_i := \{k_i\}, \; Q_i := \emptyset.$$

Крок 2: Ітеративне розширення графа

Поки $d < D$ або $L_d^{(i)} \neq \emptyset$, виконується:

1. Для множини концептів поточного рівня d:

– визначається множина публікацій $\mathcal{W}_t$, у яких зустрічається хоча б один концепт t:

$$\begin{cases} t \in L_d^{(i)} \\ \mathcal{W}_t := \{w \in \mathcal{W} \mid t \in \mathcal{T}_w\}' \end{cases}$$

де $\mathcal{T}_w$ – множина концептів, асоційованих з публікацією $w \in \mathcal{W}$;

– формується множина співвживаних концептів до концептів рівня d:

$$\mathcal{C}_d^{(i)} := \bigcup_{w \in \mathcal{W}_t} \mathcal{T}_w \setminus \bigcup_{j=0}^{d} L_j^{(i)}.$$

Це всі концепти, що зустрічаються разом із концептами {t}, але ще не були включені до графа (не зустрічаються на попередніх рівнях).

2. Для кожного $c \in \mathcal{C}_d^{(i)}$:

– обчислюється частота співвживання:

$$f_c^{(i)} := | \{ w \in \mathcal{W} \mid c \in \mathcal{T}_w \}.$$

Частота співвживання – це кількість робіт, де концепт c зустрічаються разом з хоча б одним концептом t рівня d.

Якщо $f_c^{(i)} \geq \theta$, то:

Додається вершина c до V_i на рівень d+1,

Додається ребро (k_i, c) до Q_i з вагою

$$r_i(k_i, c) := f_c^{(i)},$$

інакше переходимо розгляду до наступного концепту рівня d.

3. Формується новий рівень, а саме множина концептів рівня $(d + 1)$:

$$L_{d+1}^{(i)} := \{c \in \mathcal{C}_t \mid f_c^{(i)} \geq \theta\}.$$

4. Переходимо до опрацювання наступного рівня $d := d + 1$ (Крок 2).

Результатом роботи алгоритму є множина графів:

$$\mathcal{G} = \{G_1, G_2, \dots, G_n\},$$

де кожен граф G_i моделює семантичне оточення відповідного концепту k_i, що дозволяє здійснювати поелементну оцінку тематичної компетентності експертів на подальших етапах.

Властивості отриманих графів:

– Локальність: кожен граф G_k моделює семантичне оточення лише одного концепту.

– Ієрархічність: вершини впорядковані за рівнями $L_0, L_1, \dots, L_D$, що відповідають зростанню семантичної відстані.

– Семантична релевантність: наявність ребра між концептами означає їхню емпіричну пов'язаність у науковому дискурсі.

– Фільтрація шуму: поріг θ дозволяє виключити випадкові або нерелевантні зв'язки.

Побудовані графи $\{G_k\}_{k=1}^{n}$ використовуються для:

– оцінки тематичної компетентності експертів (через збіги з їхніми публікаційними профілями);

– формування матриці компетентностей;

– оптимізації складу експертної групи з урахуванням покриття тематики конкурсу.

Оцінка поелементної тематичної компетентності експертів

Для кожного і-го експерта-кандидата формується множина концептів, що зустрічаються в його публікаціях

$$P_j \subseteq \bigcup_{i=1}^{n} B_j^i,$$

де B_j^i – множина концептів, що асоційовані з *i*-ою публікацією експерта j.

Для кожної пари <концепт з опису предмету експертизи, експерт> (k_i, e_j) на основі побудованих графів G_i та множини P_j обчислюється оцінка компетентності S_{ij} за наступною формулою:

$$S_{ij} = \sum_{v \in V_i \cap P_j} \frac{f_j(v)}{2^{d_{k_i}(v)}},$$

де $f_j(v)$ – кількість публікацій експерта e_j, в яких зустрічається концепт v; $d_{k_i}(v)$ – номер рівня на якому зустрівся концепт v у графі G_i.

Ця формула забезпечує зважування внеску кожного концепту залежно від його семантичної віддаленості від ядра k_i, що дозволяє враховувати як прямі, так і опосередковані тематичні зв'язки.

Формування матриці компетентностей експертів:

На основі оцінок S_{ij} формується матриця:

$$M = [S_{ij}] \in \mathbb{R}^{n \times m},$$

де $i = 1, \dots, n$ – індекси концептів; $j = 1, \dots, m$ – індекси експертів.

Для забезпечення порівнюваності значень за різними концептами виконується нормалізація:

$$\hat{S}_{ij} = \frac{S_{ij}}{\max\limits_{j} S_{ij}}.$$

Отримана нормалізована матриця

$$\hat{M} = [\hat{S}_{ij}]$$

відображає, наскільки добре кожен експерт покриває окремі аспекти тематики конкурсу. Вона є основою для подальшої групової оптимізації, спрямованої на вибір експертів з комплементарними компетенціями, що забезпечують максимальне покриття концептів при мінімальному дублюванні.

Формальна постановка задачі визначення оптимального складу експертної групи:

Нехай:

$M = [S_{ij}] \in \mathbb{R}^{n \times m}$ – матриця компетентностей, де S_{ij} – оцінка компетентності експерта e_j щодо концепту k_i; n – кількість концептів; m – кількість кандидатів-експертів.

Задача полягає у виборі підмножини експертів $E^* \subseteq E$, яка:

– забезпечує максимальне покриття концептів;

– мінімізує дублювання компетенцій;

– задовольняє обмеження на розмір групи a:

$$| E^* | \leq a.$$

За наявності матриці поелементної оцінки компетентності експертів щодо кожного концепту предмету експертизи, ця задача набуває формального вигляду комбінаторної оптимізації з обмеженнями. Вибір методів її розв'язання залежить від розміру вибірки, складності цільової функції, вимог до точності, інтерпретованості та гнучкості.

Найпростішим підходом є жадібні алгоритми максимального покриття, які ітеративно додають до складу групи тих експертів, що забезпечують найбільший приріст покриття концептів. Ці методи є ефективними при невеликій кількості кандидатів і концептів, забезпечують швидке виконання та достатню прозорість, але не гарантують глобального оптимуму й можуть призводити до дублювання компетенцій.

Для задач, що потребують точного врахування обмежень (наприклад, фіксованої кількості експертів, обов'язкової присутності певних профілів, мінімального покриття кожного терміна), доцільно застосовувати методи цілочисельного програмування [11]. Вони дозволяють формалізувати задачу у вигляді лінійної або булевої моделі, яку можна розв'язати за допомогою спеціалізованих оптимізаторів. Основним обмеженням таких методів є їх обчислювальна складність, що зростає експоненційно з розміром задачі.

У випадках, коли необхідно враховувати кілька критеріїв одночасно – наприклад, тематичне покриття, збалансованість показників, рівномірність навантаження – доцільним є використання багатокритеріальних методів оптимізації. Зокрема, методи Pareto-оптимізації або зваженого агрегування дозволяють формувати множину альтернативних рішень, з яких можна обрати найкраще з урахуванням контексту. Такі підходи є гнучкими, але потребують додаткових механізмів для інтерпретації та вибору фінального рішення [12; 13].

Для задач великої розмірності, де простір можливих комбінацій експертів є надто великим для перебору, ефективними є евристичні методи, зокрема еволюційні алгоритми. Генетичні алгоритми, імітація відпалу (simulated annealing) та пошук із заборонами (tabu search) дозволяють знаходити близько-оптимальні рішення за прийнятний час. Вони добре масштабуються, легко адаптуються до складних обмежень, але мають нижчу

інтерпретованість і потребують ретельного налаштування параметрів.

Окрему групу становлять методи кластеризації, які дозволяють попередньо згрупувати експертів за тематичним профілем, а потім обрати представників з кожного кластера. Це забезпечує тематичну різноманітність і прозорість, але не гарантує оптимального покриття концептів і може потребувати додаткових евристик для вибору представників [14].

У практичному застосуванні доцільно комбінувати методи: використовувати кластеризацію для попереднього аналізу, жадібні алгоритми для швидкого старту, а еволюційні або цілочисельні моделі – для фінального уточнення складу групи. Вибір конкретного методу має базуватися на вимогах до точності, ресурсних обмеженнях, структурі матриці компетентностей та характері конкурсної тематики.

У контексті цифрової трансформації наукової експертизи запропонований підхід може бути інтегрований у системи автоматизованого добору експертів, що використовуються на рівні національних грантових агентств, редакцій наукових журналів або платформ для академічного рецензування.

Перспективним напрямом розвитку моделі є її розширення на задачі виявлення конфліктів інтересів та оцінки академічної доброчесності. Зокрема, аналіз співавторства, організаційної афіліації та цифрової взаємодії (наприклад, згадок у соціальних мережах) може бути формалізований як окремий критерій у матриці компетентностей. Це дозволить не лише підвищити об’єктивність добору, а й забезпечити відповідність етичним стандартам академічної спільноти.

Висновки

Запропонований семантико-статистичний підхід до формування експертних груп у наукових конкурсах демонструє суттєві переваги над традиційними бібліометричними методами. Модель забезпечує поелементну оцінку тематичної компетентності експертів шляхом побудови локальних графів семантичної близькості для кожного концепту конкурсної тематики. Це дозволяє враховувати як прямі, так і опосередковані зв’язки між поняттями, підвищуючи точність і змістовну релевантність оцінки.

Формалізація оцінки через зважування частоти згадування концептів та їх семантичної віддаленості забезпечує прозорість і інтерпретованість результатів. Побудована матриця компетентностей дозволяє застосовувати методи групової оптимізації для формування збалансованих експертних складів з максимальним тематичним покриттям і мінімальним дублюванням.

Модель є гнучкою щодо вибору оптимізаційної стратегії: від жадібних алгоритмів до еволюційних та цілочисельних методів, що дозволяє адаптувати її до різних масштабів і вимог. Використання концептів OpenAlex як відкритої семантичної бази забезпечує масштабованість, відтворюваність і незалежність від закритих джерел.

Подальші дослідження можуть бути спрямовані на інтеграцію динаміки наукової активності, історії рецензування, соціальних зв’язків між експертами та механізмів виявлення конфліктів інтересів, що дозволить розширити функціональність моделі в контексті академічної доброчесності та стратегічного управління експертизою.

Список літератури

1. Шостак О. Розробка підходу до формування експертних комісій щодо оцінювання складу команд виконавців високотехнологічних проєктів. *Технологічний аудит і резерви виробництва*, 2016. 4 (2). С. 20–25. URL: http://nbuv.gov.ua/UJRN/Tatrv_2016_4(2)__4.

2. Akhtar M. K. The H-index is an unreliable research metric for evaluating the publication impact of experimental scientists. *Frontiers in Research Metrics and Analytics*, 2024. 9. Article 1385080. URL: https://doi.org/10.3389/frma.2024.1385080.

3. Koltun V., Hafner D. The h-index is no longer an effective correlate of scientific reputation. *PLOS ONE*, 2021. 16 (6). e0253397. URL: https://doi.org/10.1371/journal.pone.0253397.

4. Bornmann L., Daniel H.-D. What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 2008. 64 (1). С. 45–80. URL: https://doi.org/10.1108/00220410810844150.

5. Циганок В. В., Хроленко Я. О., Доманецька І. М. Інтелектуальні засоби обробки текстів для задач організації та проведення конкурсів студентських наукових робіт. *Системи та засоби штучного інтелекту: тези доповідей Міжнародної наукової конференції «Штучний інтелект: досягнення, виклики та ризики»*. Київ: ІПШІ «Наука і освіта», 2024. С. 331–335. URL: https://essuir.sumdu.edu.ua/server/api/core/bitstreams/9c189961-9f5f-44f5-8030-d60d67f61d52/content.

6. Tang J., Zhang J., Yao L., Li J., Zhang L., Su Z. ArnetMiner: Extraction and mining of academic social networks. *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '08)*. New York: Association for Computing Machinery, 2008. С. 990–998. URL: https://doi.org/10.1145/1401890.1402008.

7. Wang F., Zhou S., Shi N. A proactive decision support system for reviewer recommendation in academia. *Expert Systems with Applications*, 2021. 169. Article 114331. URL: https://doi.org/10.1016/j.eswa.2020.114331.

8. Stelmakh I., Shah N., Singh A. PeerReview4All: Fair and accurate reviewer assignment in peer review. *Journal of Machine Learning Research*, 2021. 22(70). C. 1–55. URL: https://doi.org/10.48550/arXiv.1806.06237.

9. Balagura I., Andrushchenko V., Gorbov I. Detection of expert groups for scientific expertise. *CEUR Workshop Proceedings*, 2023. 2318. C. 271–280. URL: https://ceur-ws.org/Vol-2318/paper23.pdf.

10. Priem J., Piwowar H., Orr R. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint*, 2022. arXiv:2205.01833. URL: https://doi.org/10.48550/arXiv.2205.01833.

11. Fernández E., Rangel-Valdez N., Cruz-Reyes L., Gomez-Santillan C. Leveraging multicriteria integer programming optimization for effective team formation. *IEEE Xplore*, 2021. URL: https://ieeexplore.ieee.org/document/10531676.

12. Williams A. Insights into weighted sum sampling approaches for multi-criteria decision making problems. *arXiv preprint*, 2024. arXiv:2410.03931. URL: https://arxiv.org/abs/2410.03931.

13. Jakob W., Blume C. Pareto optimization or cascaded weighted sum: A comparison of concepts. *Algorithms*, 2014. 7(1). C. 166–185. URL: https://doi.org/10.3390/a7010166.

14. Zhou J., Zheng H., Li S., Hao Q., Zhang H., Gao W., Wang X. A knowledge-guided competitive co-evolutionary algorithm for feature selection. *Applied Sciences*, 2024. 14(11). Article 4501. URL: https://www.mdpi.com/2076-3417/14/11/4501.

Стаття надійшла до редколегії 18.11.2025

Domanetska Iryna
PhD in Technical Sciences, Associate Professor, Associate Professor of the Department of Intelligent Technologies,
https://orcid.org/0000-0002-8629-9933
Taras Shevchenko National University of Kyiv, Kyiv
Khrolenko Yaroslav
PhD Student, Department of Intelligent Decision Support Technologies,
https://orcid.org/0009-0004-0641-827X
Institute of Information Registration Problems of the National Academy of Sciences of Ukraine, Kyiv

FEATURES OF DETERMINING THE OPTIMAL STRUCTURE OF AN EXPERT GROUP BASED ON A SEMANTIC-STATISTICAL APPROACH

Abstract. *The formation of expert groups for evaluating scientific projects, grant proposals, and peer reviews is a critically important task that directly affects the quality, objectivity, and integrity of scientific evaluation. Traditional approaches to expert selection–such as ranking by the Hirsch index or classification using ASJC/MeSH codes–have several limitations: they fail to account for the semantic relevance of an expert's profile to the competition's subject area, disregard interdisciplinary links, and do not provide a fine-grained analysis of thematic coverage. In response to these challenges, this paper proposes a semantic-statistical approach to expert competence assessment, based on constructing co-occurrence graphs of terms for each concept of the competition's thematic domain using the OpenAlex concept vocabulary. The model enables element-wise evaluation of experts' thematic competence based on their publication profiles, considering both direct and indirect semantic relations. Within the study, we formalize an algorithm for building local graphs of semantic proximity for each concept and define a competence evaluation metric that incorporates both the frequency of term occurrences in publications and their semantic distance from the thematic core. Based on the obtained scores, a competence matrix is constructed, allowing the application of group optimization methods to form balanced expert panels. Several optimization strategies are examined–greedy algorithms, integer programming, and evolutionary methods–and their applicability is justified depending on task size and precision requirements. The proposed approach ensures transparency, thematic sensitivity, and flexibility in forming expert groups, prevents redundancy of competences, and avoids underrepresentation of narrowly specialized researchers. The model is scalable, reproducible, and suitable for integration into automated expert selection systems for scientific competitions, with potential extensions toward conflict-of-interest detection and academic integrity assessment.*

Keywords: thematic competence; expert group formation; semantic-statistical approach; OpenAlex Concepts; thematic coverage; term co-occurrence

References

1. Shostak, O. (2016). Development of an approach to the formation of expert commissions for evaluating the composition of teams of performers of high-tech projects. *Technological Audit and Production Reserves*, *4* (2), 20–25. URL: http://nbuv.gov.ua/UJRN/Tatrv_2016_4(2)__4.

2. Akhtar, M. K. (2024). The H-index is an unreliable research metric for evaluating the publication impact of experimental scientists. *Frontiers in Research Metrics and Analytics*, 9, Article 1385080. URL: https://doi.org/10.3389/frma.2024.1385080.

3. Koltun, V., & Hafner, D. (2021). The h-index is no longer an effective correlate of scientific reputation. *PLOS ONE*, *16*(6), e0253397. URL: https://doi.org/10.1371/journal.pone.0253397.

4. Bornmann, L., & Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64* (1), 45–80. URL: https://doi.org/10.1108/00220410810844150.

5. Tsyganok, V. V., Khrolenko, Ya. O., & Domanetska, I. M. (2024). Intelligent text processing tools for the tasks of organizing and conducting student research competitions. *Systems and Means of Artificial Intelligence: Proceedings of the International Scientific Conference "Artificial Intelligence: Achievements, Challenges and Risks"*, 331–335. URL: https://essuir.sumdu.edu.ua/server/api/core/bitstreams/9c189961-9f5f-44f5-8030-d60d67f61d52/content.

6. Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). ArnetMiner: Extraction and mining of academic social networks. *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '08)*, 990–998. URL: https://doi.org/10.1145/1401890.1402008.

7. Wang, F., Zhou, S., & Shi, N. (2021). A proactive decision support system for reviewer recommendation in academia. *Expert Systems with Applications*, 169, Article 114331. URL: https://doi.org/10.1016/j.eswa.2020.114331.

8. Stelmakh, I., Shah, N., & Singh, A. (2021). PeerReview4All: Fair and accurate reviewer assignment in peer review. *Journal of Machine Learning Research*, 22 (70), 1–55. URL: https://doi.org/10.48550/arXiv.1806.06237.

9. Balagura, I., Andrushchenko, V., & Gorbov, I. (2023). Detection of expert groups for scientific expertise. *CEUR Workshop Proceedings*, 2318, 271–280. URL: https://ceur-ws.org/Vol-2318/paper23.pdf.

10. Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *ArXiv preprint*, arXiv:2205.01833. URL: https://doi.org/10.48550/arXiv.2205.01833.

11. Fernández, E., Rangel-Valdez, N., Cruz-Reyes, L., & Gomez-Santillan, C. (2021). Leveraging multicriteria integer programming optimization for effective team formation. *IEEE Xplore*. URL: https://ieeexplore.ieee.org/document/10531676.

12. Williams, A. (2024). Insights into weighted sum sampling approaches for multi-criteria decision making problems. *ArXiv preprint*, arXiv:2410.03931. URL: https://arxiv.org/abs/2410.03931.

13. Jakob, W., & Blume, C. (2014). Pareto optimization or cascaded weighted sum: A comparison of concepts. *Algorithms*, *7* (1), 166–185. URL: https://doi.org/10.3390/a7010166.

14. Zhou, J., Zheng, H., Li, S., Hao, Q., Zhang, H., Gao, W., & Wang, X. (2024). A knowledge-guided competitive co-evolutionary algorithm for feature selection. *Applied Sciences*, *14* (11), Article 4501. URL: https://www.mdpi.com/2076-3417/14/11/4501.

Посилання на публікацію

APA Domanetska, I., & Khrolenko, Y. (2025). Features of determining the optimal structure of an expert group based on a semantic-statistical approach. *Management of Development of Complex Systems*, 64, 194–200, dx.doi.org\10.32347/2412-9933.2025.64.194-200.

ДСТУ Доманецька І. М., Хроленко Я. О. Особливості визначення оптимального складу експертної групи на основі семантико-статистичного підходу. *Управління розвитком складних систем.* Київ, 2025. № 64. С. 194 – 200, dx.doi.org\10.32347/2412-9933.2025.64.194-200.