

ОБ'ЄКТИВІЗАЦІЯ ЕКСПЕРТИЗИ НАУКОВИХ ТВОРІВ: АВТОМАТИЧНИЙ ПІДБІР ОПОНЕНТІВ ДИСЕРТАЦІЙ

Штовба С. Д.^{1*}, Петричко М. В.^{2**}

¹Донецький національний університет імені Василя Стуса, Вінниця (Україна),

²Вінницький національний технічний університет, Вінниця (Україна)

*e-mail: s.shtovba@donnu.edu.ua

**e-mail: mpetrychko@vntu.edu.ua

Сьогодні все гостріше постає проблема якісної та своєчасної експертизи наукових творів – рукописів статей, монографій, дисертацій, запитів на гранти тощо. Експертна оцінка є ключовою в сучасному управлінні науковими дослідженнями. Численні опитування вказують на значний запит наукової спільноти щодо покращення експертного оцінювання. Покращення потребує і якість експертних висновків, і процедура оцінювання. Обсяги експертної роботи колосальні, причому вони часто сконцентровані у дуже стислі часові інтервали. Наприклад, у 2023 р. на конкурс проєктів наукових досліджень і розробок для університетів та наукових установ МОНУ подано 749 запитів. Кожен запит оцінюється 5 рецензентами, що сумарно становить приблизно 3 800 експертиз з урахуванням того, що деякі рецензії відхиляються конкурсною комісією через етичні чи змістовні зауваження. За умовами конкурсу усі підготовчі, експертні та узагальнюючі процедури здійснюються протягом 2 місяців. У такому цейтноті за ручного управління експертною діяльністю досягти високої якості висновків надто складно, про що свідчать численні негативні відгуки авторів запитів на рецензії.

Ключовим етапом експертного оцінювання є підбір рецензентів відповідно до тематики твору. Однією із задач підбору рецензентів є формування ради із захисту PhD-дисертацій. Сьогодні такі ради формують у ручному режимі. Це обумовлює і корупційні ризики, і значні витрати часу на пошук та аналіз кандидатів з великими шансами пропустити кваліфікованих опонентів. Обсяг такої роботи доволі великий – щомісяця в Україні захищається орієнтовно 300 PhD-дисертацій.

Нами пропонується метод призначення команди рецензентів, який, на відміну від ізольованого підбору кандидатів, враховує здатність саме колективу рецензентів спільно оцінити роботу за усіма аспектами її тематики. Метод складається з трьох етапів. На першому етапі категоризуємо дисертацію та потенційних членів ради шляхом представлення їх тематик векторами у просторі наукових спеціальностей з Австралійсько-Новозеландської системи класифікації наук ANZSRC-2020. Вибір ANZSRC-2020 обумовлено доступністю сервісів інформаційної системи Dimensions, в якій категоризовано за ANZSRC-2020 понад 100 млн публікацій. На другому етапі розраховуємо рівень відповідності кандидатів тематиці дисертації з урахуванням спорідненості спеціальностей ANZSRC-2020. На третьому етапі підбираємо склад ради, яка відповідає тематиці дисертації з максимально

можливим ступенем. Для реалізації третього етапу запропоновано кілька алгоритмів оптимізації.

Для експериментів з підбору опонентів сформуємо датасет дисертацій [1]. Для цього скористаємось інформаційною системою NAQA.Svr. Ми зібрали інформацію за 67 дисертаціям: 17 з них відхилені міністерством через слабку відповідність тематики статей членів ради дисертації, а 50 дисертацій було захищено, включно з 17 у повторно сформованих радах. Більшість зібраних дисертацій стосується спеціальностей ІТ-галузі.

Проведемо експерименти з підбору опонентів на сформованому датасеті дисертацій. Спочатку категоризуємо ключові слова дисертації за алгоритмом категоризації ключових слів у межах спеціальностей з ANZSRC-2020. Аналогічно категоризуємо ключові слова статей членів разових рад. Для кожної ради вилучимо опонентів і спробуємо підібрати кращих з членів інших разових рад. Після вилучення опонентів отримуємо множину фрагментів разових рад, що містять голову та двох або одного рецензентів. Тепер знайдемо опонентів, додавання яких до фрагментів разових рад забезпечить їх максимально можливу відповідність тематиці дисертацій. Результати підбору опонентів порівняємо зі складом разової ради, який сформовано закладом. Ефект оцінимо середнім рівнем зміни відповідності разових рад. На рис. 1 порівнюються рівні відповідності разових рад від закладу зі знайденими за повним перебором варіантами рад. Відбувається значне покращення відповідності більшості рад. Деякі ради не покращено або рівень покращення низький. Це насамперед тим, що в датасеті розподіл дисертацій за спеціальностями є нерівномірним і датасет має малий обсяг.

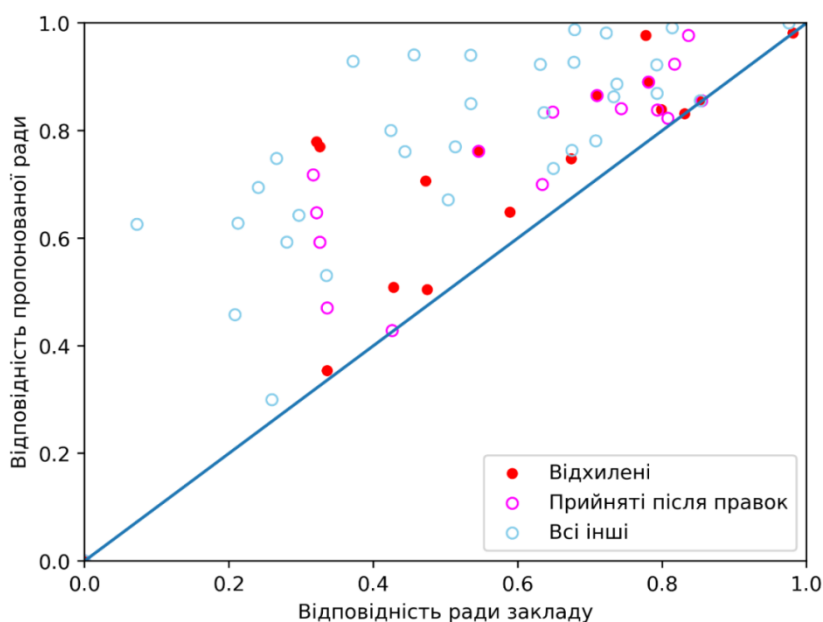


Рисунок 1. Порівняння рад від закладу із запропонованими радами

Внаслідок оптимізації вдалося покращити склад разових рад у середньому на 13–34 % залежно від типу використаного алгоритму (рис. 2). Перебір на обрізаній множині кандидатів за порогу у 0.3 однозначно невдалий. Усі інші утворюють множину Парето. Тому під час вибору алгоритму необхідно врахувати пріоритети – потрібен швидкий результат чи якісний. З рис. 2 видно, рівень покращення внаслідок переходу з жадібного пошуку на алгоритми повного перебору зростає повільно, а тривалість оптимізації зростає суттєво. Найбільш збалансованим можна вважати жадібний алгоритм підбору без елітизму. Альтернативою може бути алгоритм повного перебору з прорідженням множини кандидатів на рівні приблизно 0.25. За реальних баз даних великого обсягу тривалість оптимізації за алгоритмами повного перебору навіть на сильно прорідженій множині кандидатів може зрости занадто сильно. Тому варіації жадібних алгоритмів матимуть кращі перспективи.

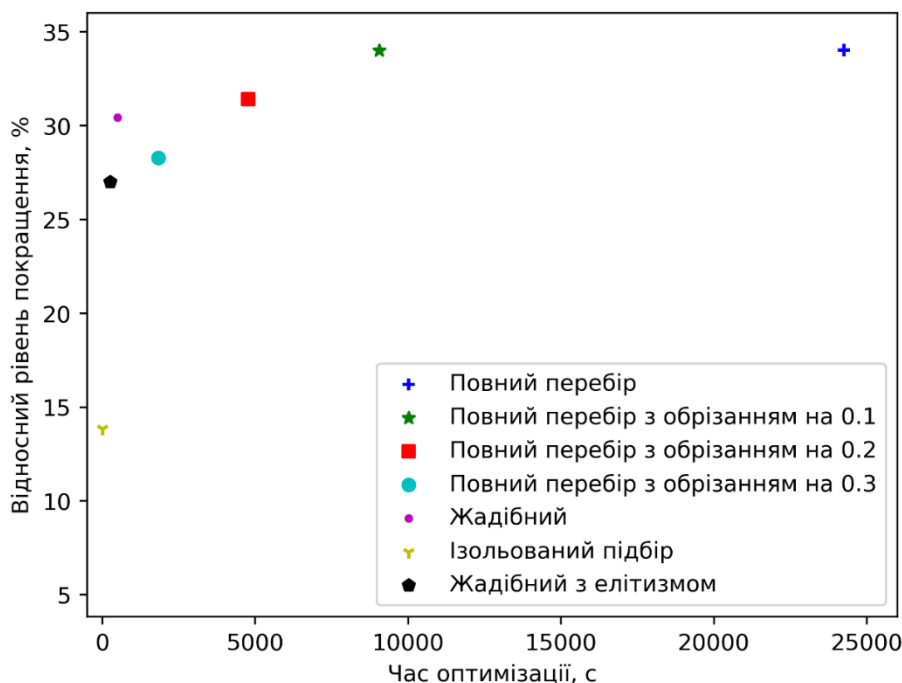


Рисунок 2. Експериментальні показники різних алгоритмів підбору опонентів

Запропонований метод може застосовуватися для покращення ефективності управління процесами призначення рецензентів для експертизи доповідей на конференціях, рецензування рукописів журнальних статей, експертизи наукових проєктів та грантових заявок, експертизи дисертацій тощо. Метод може використовуватися і в аудиторських цілях – для швидкої перевірки коректності сформованих разових рад з подальшим відбором підозрілих справ для ґрунтовної ресурсовитратної експертизи.

СПИСОК ВИКОРИСТАНИХ ПОСИЛАНЬ

1. Petrychko, M., Shtovba, S. Dataset for PhD theses reviewers assignments. *ResearchGate*. 2024. DOI: 10.13140/RG.2.2.23147.35362.