

**ІНТЕЛЕКТУАЛЬНА СИСТЕМА РЕКОМЕНДАЦІЙ
НА ОСНОВІ NLP-АНАЛІЗУ ORCID-ПУБЛІКАЦІЙ
ДЛЯ КАФЕДРИ ІІІ**

Волков К.А., Гриньова О.Є.

e-mail: kyryl.volkov@nure.ua, olena.hrynova@nure.ua

Харківський національний університет радіоелектроніки, каф. ІІІ
м. Харків, Україна

This research focuses on engineering an intelligent system designed to catalyze scientific collaboration within academic departments through NLP-based analysis of the ORCID metadata. The developed desktop solution leverages the .NET 8 ecosystem and the ML.NET framework to implement a robust recommendation engine grounded in TF-IDF vectorization and cosine similarity measures ($S > 0.7$). By autonomously identifying thematic intersections among researchers and students, the system effectively mitigates information fragmentation and facilitates the formation of an integrated, data-driven research ecosystem.

В сучасних умовах цифровізації університетської науки важливим завданням є інтеграція дослідницьких результатів та підвищення рівня міждисциплінарної взаємодії. Для сучасного університету, зокрема для кафедри (ІІІ), актуальним є створення інтелектуальних інструментів, що здатні автоматизувати процес виявлення тематичних зв'язків між дослідниками. Традиційні методи пошуку наукових партнерів через ручний моніторинг профілів виявляються неефективними в умовах експоненційного зростання обсягів наукових метаданих. Робота присвячена розробці інтелектуального застосунку для автоматизації наукової колаборації шляхом семантичного аналізу міжнародного реєстру ORCID [1].

Дана система спрямована на вирішення таких завдань:

- для здобувачів вищої освіти. Оптимізація вибору теми кваліфікаційної роботи та наукового керівника на основі об'єктивного аналізу останніх публікацій викладачів;
- планування наукової діяльності та оптимізація дослідницьких процесів кафедри штучного інтелекту. Виявлення потенційних точок наукового перетину для формування колаборацій наукових груп викладачів. Структурування інформації про актуальні напрями досліджень кафедри ІІІ.

Інтелектуальний аналіз масивів наукових публікацій дозволяє ефективно вирішити низку прикладних задач у межах кафедральної екосистеми. Зокрема, система забезпечує студенту можливість оперативно орієнтуватися в актуальній тематиці досліджень, формуючи персоналізований перелік релевантних праць за ключовими

дескрипторами, що цікавлять майбутнього фахівця. Це нівелює проблему суб'єктивності при виборі тематики кваліфікаційної роботи, оскільки система оперує об'єктивними даними публікаційної активності викладачів, що суттєво спрощує процес вибору наукового керівника, оскільки базується на об'єктивному аналізі його останніх публікацій та визначенні ступеня відповідності наукових інтересів викладача обраному науковому напрямку студента.

Для науково-педагогічного складу кафедри розроблена система стає інструментом предиктивного аналізу колаборативного потенціалу. В умовах високої динаміки сучасного наукового пошуку виникає проблема фрагментації наукового простору (інформаційної ізольованості) – ситуацій, коли дослідники одного підрозділу працюють над спорідненою проблематикою, не маючи формалізованих точок семантичного перетину. Інтелектуальний аналіз метаданих ORCID дозволяє автоматично виявляти зв'язки між різними науковими напрямками, що сприяє формуванню ефективних міждисциплінарних груп, перетворюючи розрізнені публікаційні профілі на єдиний граф наукових компетенцій кафедри. Це набуває особливої актуальності при підготовці спільних грантових заявок та плануванні комплексних науково-дослідних робіт (НДР), оскільки система надає об'єктивне підтвердження інтеграції наукових інтересів на основі ретроспективного аналізу публікаційної активності. Замість суб'єктивного пошуку партнерів, науковець отримує дата-центровану модель можливої співпраці. Такий підхід не лише оптимізує внутрішню комунікацію, а й сприяє підвищенню сумарного h-індексу кафедри та її наукового престижу в глобальному цифровому просторі, забезпечуючи перехід від стихійної до алгоритмічно обґрунтованої наукової кооперації [2].

Процес перетворення неструктурованих метаданих ORCID у релевантні рекомендації базується на багатоетапному NLP-пайплайні. Розроблене програмне забезпечення мовою C# реалізує наступний алгоритм. На першому етапі здійснюється збір даних: програма звертається до ORCID. Public API, ігноруючи блоки Employment та Education, та завантажує виключно об'єкти з секції Works. Отримані метадані включають назву публікації, анотацію, рік видання та ідентифікатори авторів. На другому етапі застосовуються методи машинного навчання. Текстові дані (Title, Abstract) проходять процес попередньої обробки, який включає нормалізацію, токенизацію та видалення стоп-слів. Далі тексти перетворюються на числові вектори ознак з використанням інструментів бібліотеки ML.NET. Для реалізації інтелектуального пошуку схожих праць було обрано метод векторизації тексту TF-IDF (Term Frequency – Inverse Document Frequency) [3]. Цей підхід дозволяє оцінити важливість слова в контексті конкретної анотації відносно всієї бази публікацій кафедри. Формалізація текстових даних дозволяє представити кожен наукову працю як точку у

багатовимірному просторі ознак. Рекомендаційна логіка ґрунтується на обчисленні косинусної схожості між векторами текстів. Для обраної праці система розраховує косинусну відстань між її вектором та векторами всіх інших праць кафедри. Роботи з коефіцієнтом схожості $S > 0.7$ пропонуються користувачу як найбільш релевантні.

Архітектурні особливості застосунку. Розробка настільного застосунку на базі .NET 8 та WPF дозволила реалізувати багатопотокову обробку даних. Поки модуль інтеграції з ORCID Public API виконує асинхронні запити до сервера, інтерфейс залишається активним для взаємодії з користувачем. Використання бібліотеки ML.NET надає можливість виконувати всі обчислення локально на машині користувача (Edge ML), що гарантує конфіденційність та високу швидкість обробки даних без необхідності постійного зв'язку з хмарними сервісами [4].

Інтерфейс програми побудовано на базі технології WPF, що забезпечує зручну візуалізацію наукового профілю кафедри, інтерактивний перегляд списку публікацій та швидкий пошук схожих досліджень. Використання настільного застосунку дозволяє забезпечити автономність роботи та інтеграцію з внутрішніми інформаційними системами університету.

Запропонована інтелектуальна система перетворює рутинний моніторинг наукових даних на алгоритмічно керований процес підтримки рішень. Пряма інтеграція з ORCID та використання методів TF-IDF дозволяють не лише скоротити час на пошук валідних джерел, а й сформувати цілісну наукову екосистему кафедри ІІІ. Це створює надійне підґрунтя для залучення молодих науковців до передових досліджень та підвищення наукового престижу університету в глобальному просторі.

Список використаних джерел:

1. ORCID. Public API [Електронний ресурс]. – 2024. – URL: <https://info.orcid.org/what-is-orcid/services/public-api/> (дата звернення: 04.03.2026).

2. Теплюк М. В., Гриньова О. Є. Підхід до створення інтелектуального помічника для наукових конференцій на основі LLM та PFE. Радіоелектроніка та молодь у XXI столітті, м. Харків. 2025. С. 63-65. URL: <https://openarchive.nure.ua/handle/document/30697> (дата звернення: 04.03.2026).

3. TF-IDF in .NET for Apache Spark Using Spark ML v2 [Електронний ресурс] / The Agile SQL Club. – 2020. – Режим доступу: <https://the.agilesql.club/2020/12/tf-idf-in-.net-for-apache-spark-using-spark-ml-v2/> (дата звернення: 04.03.2026).

4. Google AI Edge. Generate model interfaces using metadata. URL: <https://ai.google.dev/edge/litert/android/metadata/codegen> (дата звернення: 04.03.2026).