

Рисунок 2 – Приклад роботи програми. Приховування даних

Список літератури

1. Фрактали [Електронний ресурс]. – URL: <http://www.kpi.kharkov.ua/archive/microcad/2011/%95%D0%A0%D0%95%D0%94%D0%9E%D0%92%D0%98%D0%A9.pdf>

УДК 004.912

Петричко М.В., аспірант 1 курсу спеціальності
126 «Інформаційні системи та технології»
Штовба С.Д., д.т.н., професор кафедри
комп'ютерних наук та інформаційних технологій

ІДЕНТИФІКАЦІЯ НАУКОВИХ СПЕЦІАЛЬНОСТЕЙ ДОСЛІДНИКІВ НА ОСНОВІ ЇХ ІНТЕРЕСІВ В GOOGLE SCHOLAR

Вінницький національний технічний університет, м. Вінниця
Донецький національний університет імені Василя Стуса, м. Вінниця

Сьогодні професійні спільноти людей взаємодіють в різноманітних онлайн-мережах. Не виключенням є і спільнота дослідників. Найбільшою онлайн-мережею дослідників є Google Scholar. В ній, зокрема, у відкритому доступі є понад 50 тисяч профілів українських дослідників. Такий величезний ресурс виглядає привабливим для розробки технологій аналітичного опрацювання накопиченої в ньому інформації з метою ідентифікації лідерів – статей, науковців, університетів, журналів, виявлення тенденцій наукових

досліджень, кластеризації науковців, підбору партнерів для спільних проєктів, опонентів дисертацій, рецензентів рукописів тощо. Найчастіше з профілів науковців в Google Scholar використовують дані щодо цитованості. Її використовують як одне із джерел інформації для рейтингування університетів у Webometrics. Також створено кілька інформаційних систем на базі Google Scholar, найбільш відомими серед яких є Publish or Perish та Scholarometer [1]. Багато досліджень, зокрема [2, 3], стосуються перевірки відповідності між цитованістю в Google Scholar та у наукометричних системах Scopus, Web of Science, Dimensions та іншими, які наповнюються виключно за мета-даними з видавництва. Окрім списку публікацій та їх цитування в профілі науковця міститься і інша інформація. Зокрема, науковець в профілі вказує свої інтереси, і робить він це на власний розсуд, обираючи слова у довільний спосіб. Google Scholar дозволяє здійснити пошук науковців за тим чи іншим інтересом. Але видачі формуються за буквальним співпадінням. Наприклад, видачі для *fuzzy set* та *fuzzy sets* будуть різними. Google Scholar не враховує і сукупність інтересів користувача, тобто пошук за кожним інтересом здійснюється незалежно та ізольовано. Таким чином, пошукові та аналітичні сервіси за велетенським масивом профілів науковців в Google Scholar є досить примітивними.

Метою роботи є ідентифікація наукових спеціальностей дослідників на основі їх інтересів в Google Scholar. Методи опрацювання інтересів з профілів науковців в Google Scholar є достатньо мало дослідженими. Нами виявлено лише 2 релевантні публікації [4, 5]. На відміну від тих робіт, ми прагнемо категоризувати науковців в межах деякої класифікації наук, тобто розподілити їх за науковими спеціальностями. Автоматична категоризація науковців здійснюється зазвичай в результаті узагальнення тематик їх публікацій. Приклади таких розробок наведені в [6-8]. Використання такого підходу передбачає наявність достатньої кількості статей науковця з виділеними ключовими словами. При цьому, не враховується, що співавторами статті можуть бути кілька науковців, на кожного з якого припадає якась підмножина з усього списку ключових слів. Причому, з десятка ключових слів статті внесок деякого співавтора може відображати лише одне ключове слово. Крім того, науковець, особливо на початку кар'єри, може і не мати достатньої кількості статей для достовірної категоризації. Але він може самостійно задати у профілі набір ключових слів, який описує його дослідження. Науковець також може і змінити напрямок своєї діяльності, наприклад, перейшовши працювати у іншу лабораторію чи на інший проєкт. Але його продовжуватимуть категоризувати за минулими публікаціями та, відповідно, за застарілими інтересами. В зв'язку з цим виникла зацікавленість у категоризації дослідника на основі його інтересів, які дослідник власноруч сформулював на поточний момент, тобто на основі початкової інформації, що вільна від наведених вище недоліків.

Вважатимемо відомими:

$W = (w_1, w_2, \dots, w_n)$ – список ключових слів, якими науковець у своєму профілі в Google Scholar на власний розсуд описав свої інтереси;

$T = (t_1, t_2, \dots, t_m)$ – перелік можливих тем у формі списку наукових спеціальностей за деякою класифікацією наук;

D_1, D_2, \dots, D_m – тематичні колекції розмічених текстів, кожна з яких містить лише публікації, які віднесено до тем t_1, t_2, \dots, t_m , відповідно;

$B = D_1 \cup D_2 \cup \dots \cup D_m$ – загальна колекція розмічених текстів, тобто множина публікацій, кожна з яких віднесена до однієї або декількох тем з множини T .

Задача полягає у знаходженні тем з T , яким відповідає сукупність інтересів W . Будемо вказувати не лише сам факт належності, але і ступінь належності:

$$\tilde{W} = \left(\frac{\mu_W(t_1)}{t_1}, \frac{\mu_W(t_2)}{t_2}, \dots, \frac{\mu_W(t_m)}{t_m} \right),$$

де $\mu_W(t_p) \in [0, 1]$ – ступінь належності сукупності інтересів W до спеціальності t_p , $p = \overline{1, m}$.

Для отримання списку ключових слів науковця скористаємося його профілем в Google Scholar. Для прикладу на рис. 1 наведено профіль науковця з такими ключовими словами, які виділено синім кольором. Для цього науковця $w_1 = \text{"Data Science"}$, $w_2 = \text{"Information Technology"}$, $w_3 = \text{"Scientometrics"}$, $w_4 = \text{"Fuzzy Logic"}$ та $w_5 = \text{"Operations Research"}$.

Задачу пропонується вирішити з використанням сервісів інформаційної системи Dimensions, яка індексує понад 110 млн публікацій з різних галузей знань. Усі публікації в Dimensions категоризовано за дворівневим варіантом Австралійсько-Новозеландського стандарту ANZSRC (Australian and New Zealand Standard Research Classification). Запит до інформаційної системи Dimensions сформуємо окремо за кожним елементом з множини W . Результат запиту містить кількість публікацій за кожною спеціальністю та за кожною галуззю, в назві або в рефераті яких фігурує пошуковий вираз. Далі використовуючи розроблений нами алгоритм відберемо наукові спеціальності, які мають найбільшу частку публікацій з вибіленим набором ключових слів. При цьому, відфільтруємо стоп-слова та ключові слова, видачі за якими є статистично недостовірними.

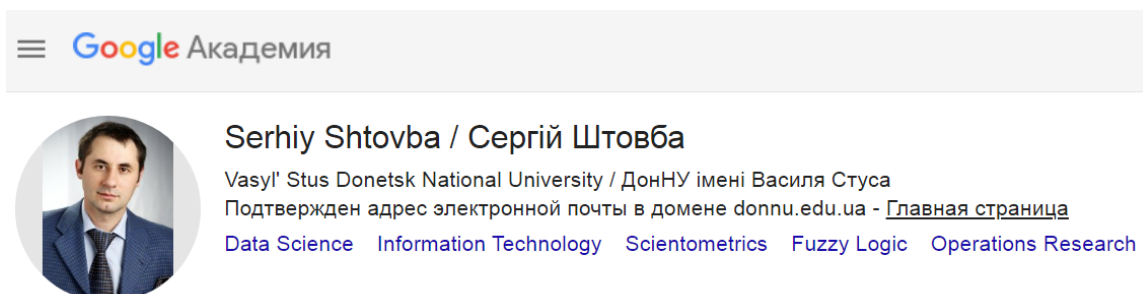


Рисунок 1 - Приклад профілю науковця в Google Scholar

Список використаної літератури

1. López-Cózar, E. D., Orduña-Malea, E., Martín-Martín, A., & Ayllón, J. M. Google Scholar: the big data bibliographic tool. In Research analytics: boosting university productivity and

- competitiveness through scientometrics (edited by Francisco J. Cantú-Ortiz) Boca Raton, FL : Taylor & Francis, 2018. P. 59-80.
2. Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & López-Cózar, E. D.. Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics*, Vol. 126, pages871–906(2021).
 3. Harzing, A.-W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787–804.
 4. Rahdari, B., Brusilovsky, P., Babichenko, D., Littleton, E. B., Patel, R., Fawcett, J., & Blum, Z. (2020). 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.
 5. Saad-Falcon, J., Shaikh, O., Wang, Z. J., Wright, A. P., Richardson, S., & Chau, D. H. (2020). PeopleMap: Visualization Tool for Mapping Out Researchers using Natural Language Processing. *arXiv preprint arXiv:2006.06105*.
 6. Rosen-Zvi, M., Griffiths, T., Steyvers, M., Smith, P., The author-topic model for authors and documents. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence* (2004), AUAI Press, 487--494.
 7. D. Blei, A. Ng. and M. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993-1022. 2003.
 8. Jian Jin, Qian Geng, Haikun Mou, Chong Chen. (2018) Author–Subject–Topic model for Reviewer Recommendation, *JIS-Journal of Information Science*, SAGE, 1-16.

УДК 004.05

Присіч А.В., студент 4 курсу
спеціальності 122 «Комп'ютерні науки»
Римар П.В., старший викладач кафедри
комп'ютерних наук та
інформаційних технологій

РОЗРОБКА СЕРВЕРНОЇ ЧАСТИНИ ДЛЯ НАВЧАЛЬНОЇ СИСТЕМИ

Донецький національний університет імені Василя Стуса, м. Вінниця

Клієнт-серверна архітектура постає домінуючою технологією при створенні розподілених мережових програм, забезпечує взаємодію та обмін даними між розробленими програмними продуктами. Сервери і клієнти такої архітектури є незалежними один від одного і функціонують паралельно. Відсутність жорсткої прив'язки між клієнтами та серверами обумовлює можливість ситуації, коли один сервер одночасно обробляє запити від різних клієнтів чи клієнт звертається по чергові до різних серверів.

До переваг клієнт-серверної архітектури відносять зменшення навантаження на клієнтські місця, що, в свою чергу, призводить до зменшення вимог до апаратно-програмного забезпечення 4 клієнтів та зниження вартісних показників системи в цілому.