

АВТОМАТИЧНА КЛАСИФІКАЦІЯ ПРОФІЛІВ В СИСТЕМІ GOOGLE SCHOLAR

Вінницький національний технічний університет

Анотація

На базі наукометричної мережі Google Scholar проведено автоматичну класифікацію профілів науковців. Описано основні алгоритми автоматичної класифікації. Проведено аналіз отриманих результатів.

Ключові слова: класифікація, наукометрична мережа, бібліографічна база даних, Google Scholar

Abstract

Google Scholar profiles automated classification are described. Main algorithms of the profile classification are provided. Classification results are shown.

Keywords: classification, research network, bibliographic database system, Google Scholar

Вступ

Останнім часом зріз інтерес до обробки різноманітної інформації, пов'язаною з науковою діяльністю. Для цього використовують різноманітні інформаційні ресурси – від спеціалізованих наукометричних бази з ретельно відібраним контентом, таких як Scopus та Web of Science, до різноманітних наукових соціальних мереж типу ResearchGate, наповнювання яких здійснюється переважно користувачами. Проміжну ланку займає Google Scholar, в якій контент виділяється з інтернету пошуковою системою Google за спеціальним алгоритмом виявлення академічних текстів. Приєднання ж цих академічних текстів до персональних профілів авторів управляється самими науковцями.

Google Scholar є безкоштовною системою, тому більшість науковців створили в ній персональні сторінки – профілі для систематизованої презентації своїх наукових публікацій, відслідковування цитувань своїх робіт, оголошення про свої наукові інтереси тощо. Останнім часом, все частіше використовують початкову інформацію з профілів науковців в Google Scholar для виявлення різноманітних закономірностей та ефектів розвитку науки. Незважаючи, що наповненість профілів сильно залежить від порядності та акуратності науковця, експериментальні дослідження підтверджують прийнятний рівень достовірності висновків, зроблених за цими даними.

Під час застосування інформаційних технологій обробки профілів науковців в Google Scholar виникають певні труднощі. Одна із яких полягає в тому, що профілі в Google Scholar можуть належати не окремому науковцю, а цілому колективу – кафедрі, лабораторії, бібліотеці тощо. Профілі можуть мати також журнали чи проекти. Крім того, для створення рекомендаційних систем різного призначення доцільно вміти автоматично розрізняти тих науковців, хто активно працює, і тих, чия активна наукова діяльність була в минулому. Метою роботи є створення моделей та відповідної інформаційної технології, яка за даними профіля в Google Scholar, автоматично віднесе його власника до перерахованих вище класів.

1 Збір початкових даних

Початковою інформацією є дані з профільної сторінки у Google Scholar. Приклади профільних сторінок наведено на рис. 1 та 2.

The screenshot shows the profile of Serhiy Shtovba (Штовба Сергей Дмитриевич / Serhiy Shtovba) at Vinnytsia National Technical University. The profile includes a bio, a list of publications with citation counts and years, a citation statistics table, a bar chart of citations from 2009 to 2016, and a list of co-authors.

Штовба Сергей Дмитриевич / Serhiy Shtovba ПОДПИСКА ОФОРМЛЕНА

Винницький національний технічний університет / Vinnytsia National Technical University
Подтвержден адрес электронной почты в домене vntu.edu.ua - [Главная страница](#)
искусственный интеллект информационные техноло... Fuzzy Logic Data Science

НАЗВАНИЕ	ПРОЦИТИРОВАНО	ГОД
<input type="checkbox"/> Проектирование нечетких систем средствами MATLAB СД Штовба М.: Горячая линия–Телеком	1112 *	2007
<input type="checkbox"/> Введение в теорию нечетких множеств и нечеткую логику [Электронный ресурс] СД Штовба Режим доступа: http://matlab.exponenta.ru/fuzzylogic/book1/index.php	553 *	2001
<input type="checkbox"/> Муравьиные алгоритмы СД Штовба Exponenta Pro. Математика в приложениях 4, 70-75	205	2003
<input type="checkbox"/> Ant algorithms: theory and applications SD Shtovba Programming and Computer Software 31 (4), 167-178	114	2005
<input type="checkbox"/> Нечеткая надежность алгоритмических процессов АП Ротштейн, СД Штовба Винница: Континент-ПРИМ	107	1997
<input type="checkbox"/> Нечеткий многокритериальный анализ вариантов с применением парных сравнений АП Ротштейн, СД Штовба Известия РАН. Теория и системы управления, 150-154	93	2001
<input type="checkbox"/> Обеспечение точности и прозрачности нечеткой модели Мамдани при обучении по экспериментальным данным	67	2007

Процитировано [ПРОСМОТРЕТЬ ВСЕ](#)

	Все	Начиная с 2013 г.
Статистика цитирования	3080	1738
h-индекс	18	14
i10-индекс	36	21

2009 2010 2011 2012 2013 2014 2015 2016 2017 2018

Соавторы [ИЗМЕНИТЬ](#)

- Ольга Панкевич - Olga Pankevych
Винницький національний технічний університет
- Alexander Rothstein
- Олег Козачко
доцент інженерної графіки, Вінницький національний технічний університет
- Анастасія Нагорна /Anastasiia Nagorna
Винницький національний технічний університет

Рисунок 1 – Профіль у Google Scholar першого автора статті

The screenshot shows the profile of Roman Tylets (Тилець Роман Олексійович / Roman Tylets) at Vinnytsia National Technical University. The profile includes a bio, a list of publications, and a list of co-authors.

Тилець Роман Олексійович / Roman Tylets ПОДПИСАТЬСЯ

Vinnytsia national technical university, master
Подтвержден адрес электронной почты в домене inmad.vntu.edu.ua - [Главная страница](#)
artificial intelligence decision theory fuzzy logic

НАЗВАНИЕ	ПРОЦИТИРОВАНО	ГОД
<input type="checkbox"/> ІДЕНТИФІКАЦІЯ ЗАЛЕЖНОСТЕЙ ЗА ДОПОМОГОЮ ГІБРИДНОЇ НЕЧІТКОЇ БАЗИ ЗНАТЬ З РІЗНОРІДНИМИ ПРАВИЛАМИ СД Штовба, АВ Нагорна, ТР О. XII Міжнародна конференція Контроль і управління в складних системах (КУСС-2014)		2014
<input type="checkbox"/> ВИКОРИСТАННЯ НЕЧІТКИХ БАЗ ЗНАТЬ З ПРАВИЛАМИ РІЗНИХ ТИПІВ ДЛЯ ІДЕНТИФІКАЦІЯ БАГАТОФАКТОРНИХ ЗАЛЕЖНОСТЕЙ СД ШТОВБА, АВ ГАЛУЩАК, РО ТИЛЕЦЬ InfoCom 2015, 201520		

Соавторы

- Штовба Сергей Дмитриевич / Serhiy...
Винницький національний технічний університет
- Анастасія Нагорна /Anastasiia Nagorna...
Винницький національний технічний університет

Рисунок 2 – Профіль у Google Scholar другого автора статті

Профіль містить у собі ім'я власника, місце роботи, список наукових інтересів, публікації та інформація про них (рік видання, назва журналу, індекс цитування та ін.), показники цитування власника тощо (рис. 3).

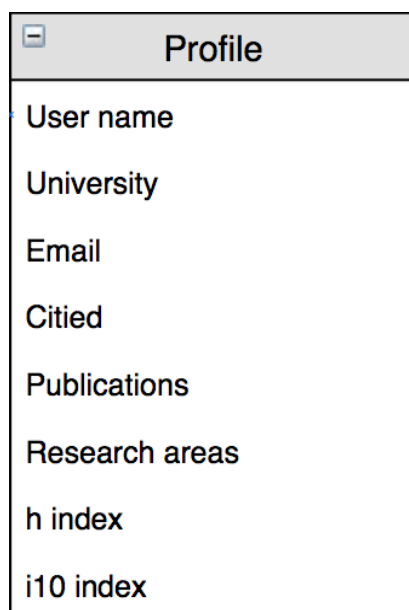


Рисунок 3 – Інформаційна модель профілю Google Scholar

Для автоматичного збору інформації з профілів у Google Scholar розроблено окремий модуль Java програми

2 Правила класифікації

Класифікація профілю проводилася в два етапи:

1. Ідентифікація профілів що не є профілями вчених. До таких профілів належать профілі кафедр, інститутів, журналів та подібні.
2. Верифікація неактивних профілів вчених.

Ідентифікація профілів що не є профілями вчених:

1. Профілі в назві яких присутні ключові слова: кафедра, інститут, журнал, вісник, бібліотека, факультет, лабораторія, ISSN, НДР, секція та їхні синоніми. Також автоматично проводиться переклад ключових слів російською та англійською мовами. Список ключових слів доповнюється в ручному режимі.

Верифікація неактивних профілів вчених:

1. Аналіз публікацій профілю. Профілі які не містять публікації за останні N років. N обрано 5.
2. Профілі які не містять хоча б 1 науковий інтерес.
3. Аналіз імені профіля. Якщо в імені присутні дати життя/смерті.
4. Профілі які не містять публікацій.
5. Профілі які містять публікації яким більше 60 років.

3 Візуалізація отриманих даних

Для прикладу проведено класифікації профілей Вінницького національного технічного університету, які містять email адреси з домену vntu.edu.ua. Статистика наведена на рис. 4. та рис. 5.

Загальна кількість профілей – 541.

Кількість профілей активних вчених – 227 (42%).

Кількість профілей не активних вчених – 308 (57%).

Кількість неперсональних профілей – 6 (1%), див. табл. 1.

Кількість профілів із заповненими науковими напрямками – 483 (89%).

Кількість профілей з незаповненими науковими напрямками – 58 (11%).

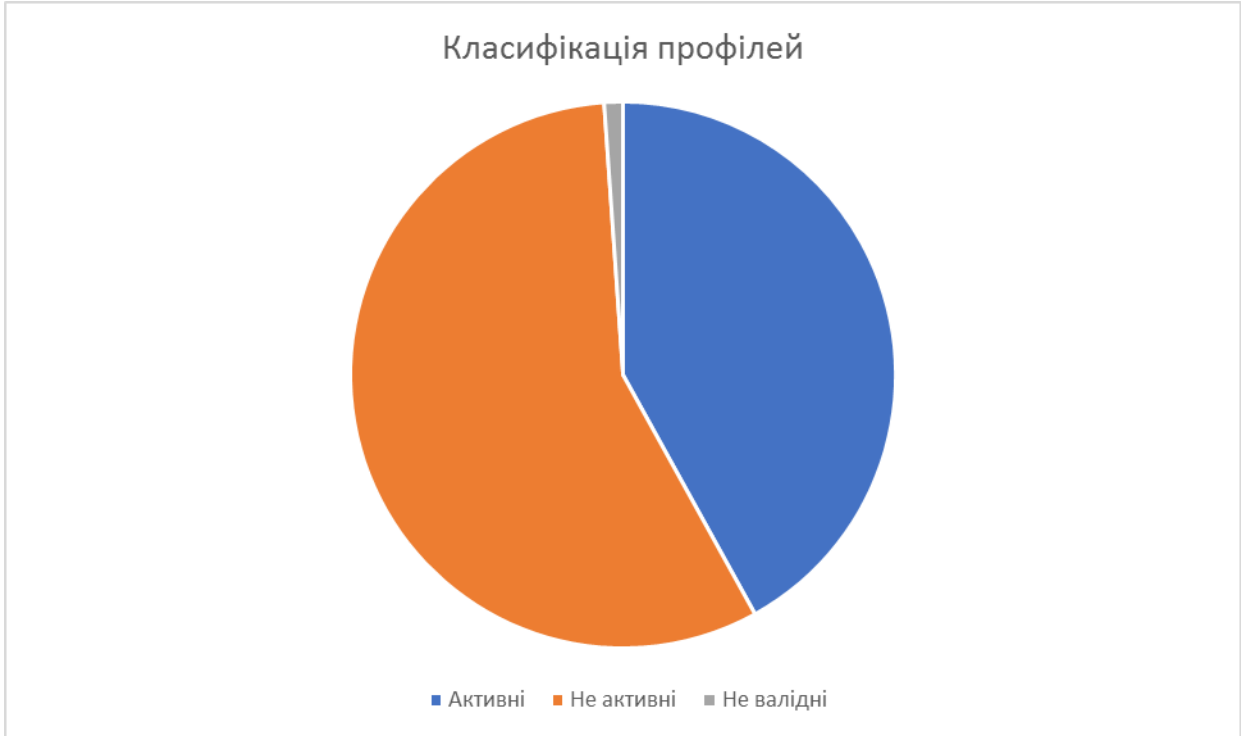


Рисунок 4 – Класифікація профілей зареєстрованих в домені vntu.edu.ua

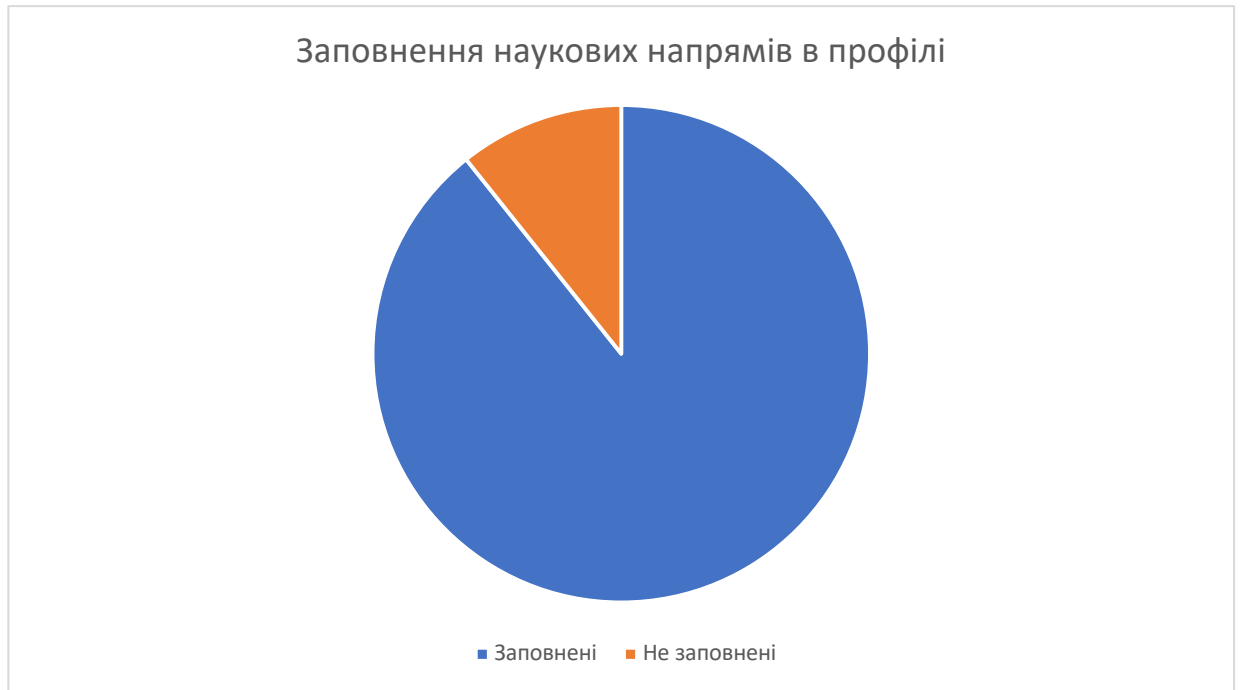


Рисунок 5 – Профілі з заповненим та незаповненими науковими напрямками

Таблиця 1 – Профілі з Вінницького національного технічного університету, що не є сторінками вчених

Назва	Посилання
Кафедра обчислювальної техніки / Department of Computer Facilities	https://scholar.google.com.ua/citations?user=7YV6C2AAAAAJ&hl=en&oe=ASCII
Кафедра метрології та промислової автоматики	https://scholar.google.com.ua/citations?user=aHeNrFgAAAAAJ&hl=en&oe=ASCII
Кафедра комп'ютерних наук ВНТУ	https://scholar.google.com.ua/citations?user=9jPLOoQAAAAAJ&hl=en&oe=ASCII
Вісник Вінницького політехнічного інституту	https://scholar.google.com.ua/citations?user=-3A06EsAAAAAJ&hl=en&oe=ASCII
кафедра мовознавства	https://scholar.google.com.ua/citations?user=F66zP5cAAAAAJ&hl=en&oe=ASCII
Вісник машинобудування та транспорту	https://scholar.google.com.ua/citations?user=La4Wy3oAAAAAJ&hl=en&oe=ASCII

Висновки

Класифікація профілей Google Scholar показала що існує багато профілей які не підтримуються в належному стані власниками, не заповнені тощо. Дана класифікація в подальшому буде використана для побудови мапи наук університету.

Штовба Сергій Дмитрович – д.т.н., професор, професор кафедри комп'ютерних систем управління, e-mail: shtovba@vntu.edu.ua;

Тилець Роман Олексійович - аспірант кафедри комп'ютерних систем управління, e-mail: roman.tylets@gmail.com;

Shtovba Serhiy – Dr. Sc. (Eng.), Professor, Professor of Computer Control Systems Department, Vinnytsia National Technical University, Vinnytsia. e-mail: shtovba@vntu.edu.ua.

Tylets Roman – PhD student of Computer Control Systems Department, Vinnytsia National Technical University, Vinnytsia, e-mail: roman.tylets@gmail.com;