

**ІНФОРМАЦІЙНА СИСТЕМА ІЄРАРХІЧНОЇ
КАТЕГОРИЗАЦІЇ НАУКОВЦІВ НА ОСНОВІ
ВІДКРИТИХ ДАНИХ З GOOGLE SCHOLAR.
ЧАСТИНА 1: ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ
ДАНИХ**

Виконав студент групи 2АКІТ-18м

спеціальності

«Автоматизація та комп'ютерно-інтегровані технології»

Петричко М.В.

Керівник: д.т.н., професор Штовба С.Д.

АКТУАЛЬНІСТЬ

label:fuzzy_sets



Zeshui Xu (徐泽水)

Academician of IASCYS, FIEEE, FIFSA, FIET, FBOS, FRSA; Sichuan University
Підтверджена електронна адреса в scu.edu.cn

[Decision Making](#) [Information Fusion](#) [Fuzzy Sets](#) [Computational Intelligence](#)
[Bibliometrics](#)

Цитовано в 53565 джерелах



János Fodor

Óbuda University
Підтверджена електронна адреса в uni-obuda.hu

[Computational intelligence](#) [fuzzy sets](#) [preference modelling](#)

Цитовано в 46920 джерелах



Enrique Herrera-Viedma

Professor of Computer Science and Artificial Intelligence, Spain, University of Granada
Підтверджена електронна адреса в decsai.ugr.es

[Fuzzy sets](#) [fuzzy decision making](#) [computing with words](#)
[multiple criteria decision making](#) [consensus](#)

Цитовано в 32635 джерелах



Lawrence Hall

Professor of Computer Science and Engineering, University of South Florida
Підтверджена електронна адреса в mail.usf.edu

[artificial intelligence](#) [pattern recognition](#) [data mining](#) [fuzzy sets](#)

Цитовано в 25236 джерелах



Peter Fisher

Professor of Geographical Information, GIS & Remote Sensing Group, University of Leicester ...
Підтверджена електронна адреса в le.ac.uk

[Geographical Information Science](#) [Spatial Analysis](#) [Uncertainty](#) [Fuzzy Sets](#)
[Remote Sensing](#)

Цитовано в 17404 джерелах

label:fuzzy_set



Hongying Zhang

Xi'an jiaotong university
Підтверджена електронна адреса в mail.xjtu.edu.cn

[rough set](#) [fuzzy set](#)

Цитовано в 38944 джерелах



Rajkumar Sharma

Assistant Professor of Information Technology, LNCT Bhopal
Підтверджена електронна адреса в lnctgroup.co.in

[Machine Learning](#) [Data Mining](#) [Rough Set Theory](#) [Fuzzy Set](#) [Big Data & Hadoop](#)

Цитовано в 9145 джерелах



Weihua Xu

Professor of Congqing University of Technology
Підтверджена електронна адреса в cqu.edu.cn

[Rough set](#) [Fuzzy set](#) [Artificial intelligence](#)

Цитовано в 2608 джерелах



Chongfu Huang

Beijing Normal University
Підтверджена електронна адреса в bnu.edu.cn

[Risk Analysis](#) [Fuzzy Set](#) [Natural Disaster](#)

Цитовано в 2211 джерелах



Qingguo Li (李庆国)

Hunan University
Підтверджена електронна адреса в hnu.edu.cn

[Domain Theory](#) [Fuzzy Set](#) [Rough Set](#) [Three-way Decision](#)

Цитовано в 1474 джерелах

ОБ'ЄКТ, ПРЕДМЕТ ТА МЕТА ДОСЛІДЖЕННЯ

- **Об'єкт дослідження** – інформаційні системи підтримки наукової діяльності.
- **Предмет дослідження** - моделі та алгоритми ієрархічної категоризації науковців.
- **Метою комплексної роботи** є розробка інформаційної системи ієрархічної категоризації науковців за відкритими даними з Google Scholar.
- **Мета роботи** полягає у створенні інформаційної системи для категоризації науковців за спеціальностями та галузями науки.

ЗАВДАННЯ ДОСЛІДЖЕННЯ

- розробити модель представлення науковців у просторі галузей та спеціальностей за деякої специфікації наук;
- дослідити різноманітні системи класифікації наук та обґрунтувати вибір найбільш придатної для вирішення задачі категоризації;
- дослідити схожість наукових спеціальностей та розробити модель їх оцінювання;
- покращити алгоритм категоризації науковців за спеціальностями який враховує взаємодію як інтересів науковців, так і наукових спеціальностей;
- реалізувати програмно запропоновані моделі та алгоритми.

ІНФОРМАЦІЙНІ СИСТЕМИ ПІДТРИМКИ НАУКОВОЇ ДІЯЛЬНОСТІ

Scopus®

 Springer

 Dimensions

Бібліометрика української науки

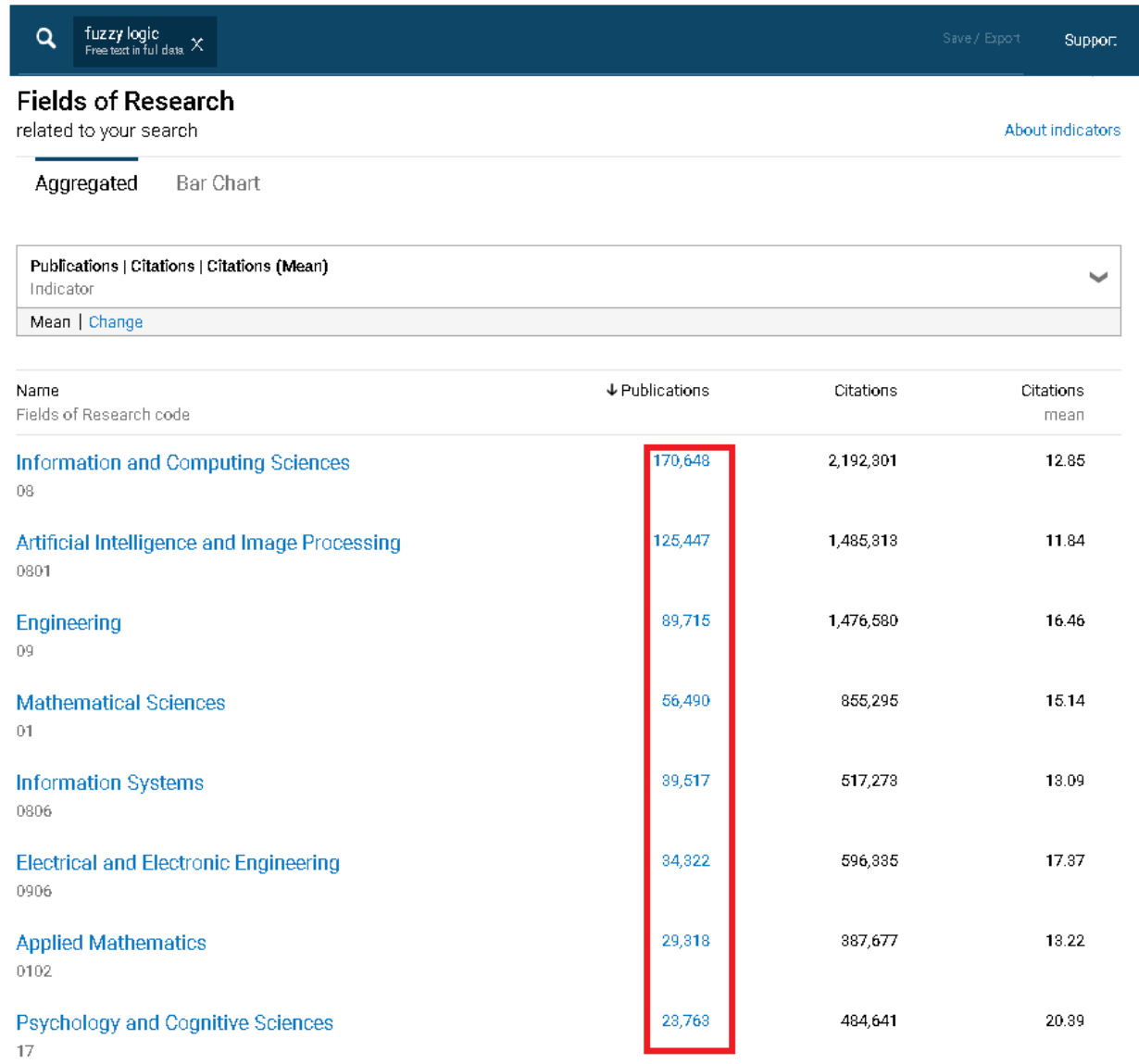
Бібліометричний профіль вченого - декларація про наукову діяльність



WEB OF SCIENCE

 **Google**
Scholar

ІДЕЯ МЕТОДУ



ОПИС АЛГОРИТМУ

Метод категоризації складається з таких процедур:

- **розширення множини інтересів науковця парними запитами, наприклад, *fuzzy logic & machine learning*;**
- підрахунок частоти входження інтересу науковця до спеціальностей науки;
- нехтування непопулярних спеціальностей науки;
- агрегування частот входження кожного інтересу;
- **редукція спеціальностей на основі індексу схожості спеціальностей.**



СХОЖІСТЬ НАУКОВИХ СПЕЦІАЛЬНОСТЕЙ У СИСТЕМІ ANZSRC

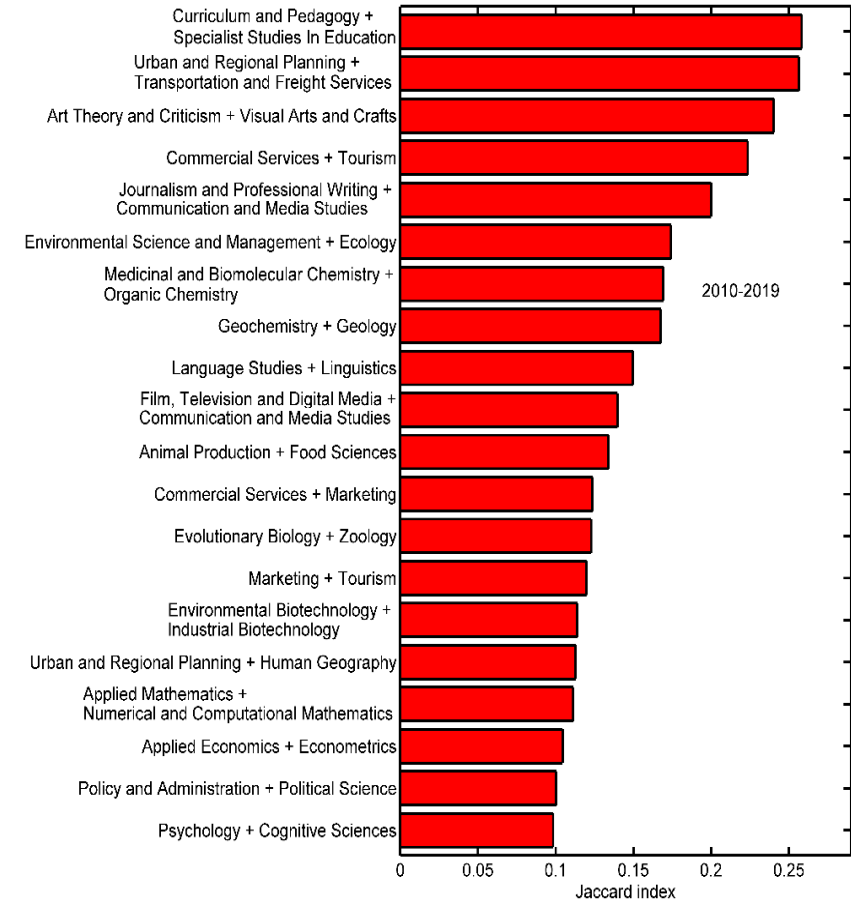
Модель схожості спеціальностей (індекс Жаккара):

$$S(A, B) = \frac{N_{A \cap B}}{N_A + N_B - N_{A \cap B}}$$

N_A – кількість публікацій з спеціальністю A

N_B – кількість публікацій з спеціальністю B

$N_{A \cap B}$ – кількість публікацій одночасно з спеціальностями A та B

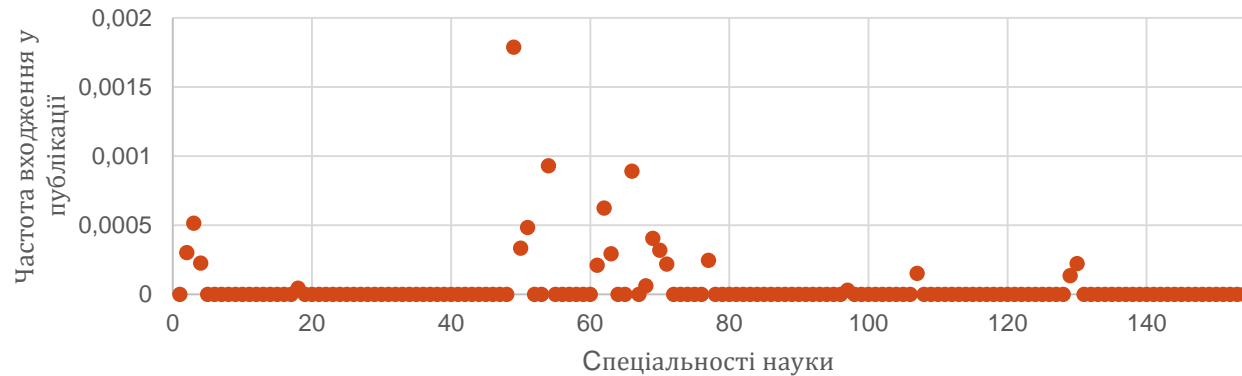


ՏՄՏԵՄԱ ԽԱՅԿ ԱՆՀՏՐԸ

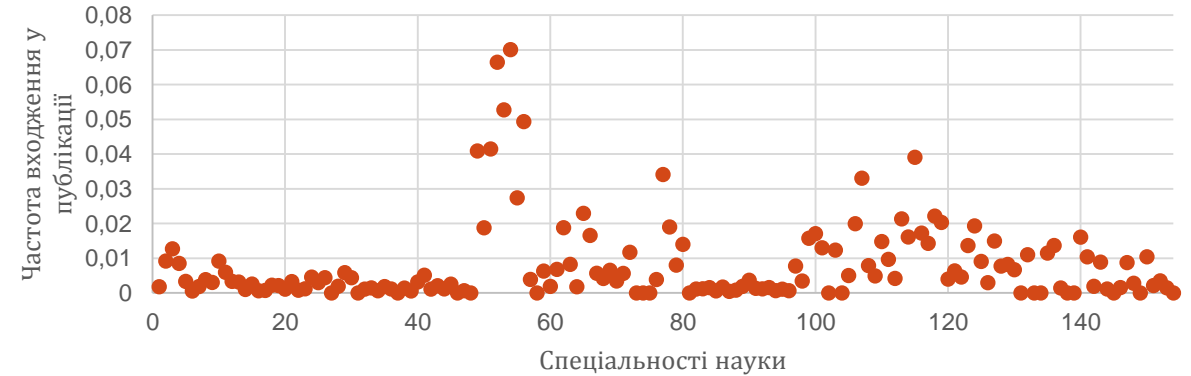
- **01 Mathematical Sciences**
 - 0101 Pure Mathematics
 - 0102 Applied Mathematics
 - 0103 Numerical and Computational Mathematics
 - 0104 Statistics
 - 0105 Mathematical Physics
- **02 Physical Sciences**
- **03 Chemical Sciences**
- **04 Earth Sciences**
- **05 Environmental Sciences**
- **06 Biological Sciences**
- **07 Agricultural and Veterinary Sciences**
- **08 Information and Computing Sciences**
- **09 Engineering**
- **10 Technology**
- **11 Medical and Health Sciences**
- **12 Built Environment and Design**
- **13 Education**
- **14 Economics**
- **15 Commerce, Management, Tourism and Services**
- **16 Studies in Human Society**
- **17 Psychology and Cognitive Sciences**
- **18 Law and Legal Studies**
- **19 Studies in Creative Arts and Writing**
- **20 Language, Communication and Culture**
- **21 History and Archaeology**
- **22 Philosophy and Religious Studies**

ПРИКЛАД КАТЕГОРИЗАЦІЇ (І ЕТАП)

artificial intelligent



information technology



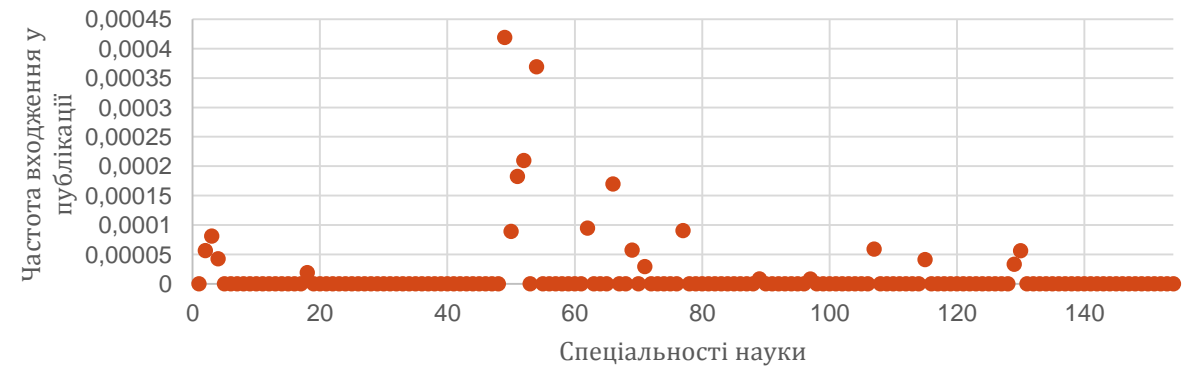
Roman Tkachenko

Dr.Sc., Prof., [Lviv Polytechnic National University](#)

Підтверджена електронна адреса в [Irpri.ua](#) - [Домашня сторінка](#)

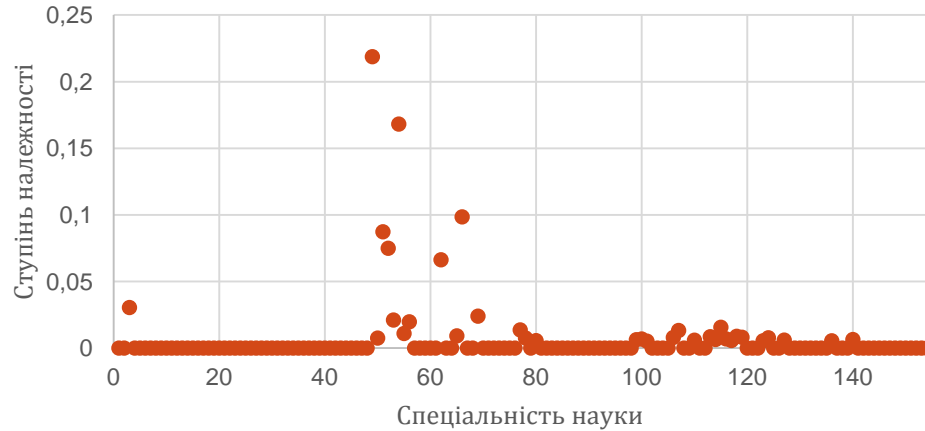
[artificial intelligent](#) [information technology](#)

artificial intelligent & information technology

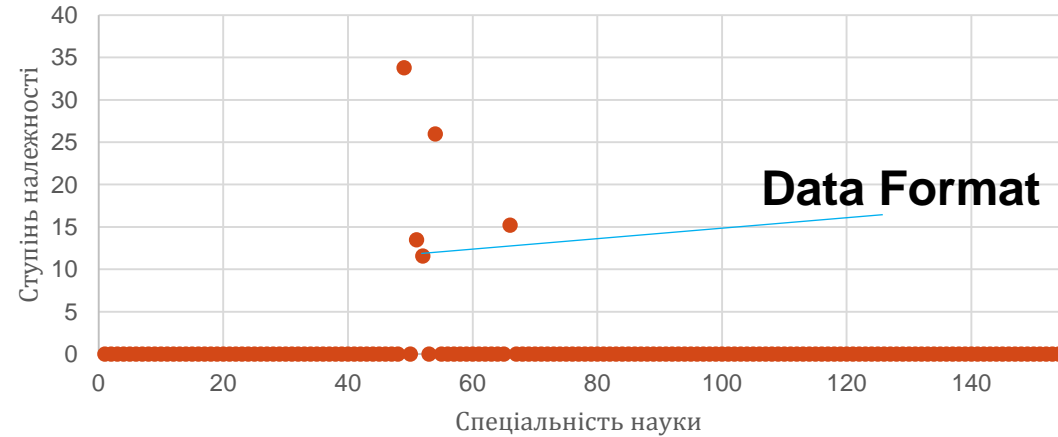


ПРИКЛАД КАТЕГОРИЗАЦІЇ

Після агрегування за інтересами (II)



Після нехтування непопулярних спеціальностей (III)



Редукція спеціальностей з перерахунком впливу (IV)



Результат категоризації:

- **08 Information and Computing Science – 83.11**
 - 0801 Artificial Intelligence and Image Processing – **37.98**
 - 0806 Information Systems – **29.82**
 - 0803 Computer Software – **15.27**
- **09 Engineering – 16.91**
 - 0910 Manufacturing Engineering – **16.91**

АПРОБАЦІЯ РЕЗУЛЬТАТІВ

Науково-технічні конференції

- Петричко М. В. Статистичний підхід до оцінювання подібності наукових спеціальностей в системі Dimensions [Електронний ресурс] / М. В. Петричко, С. Д. Штовба // ВНТУ. – 2019. – Режим доступу до ресурсу: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2019/paper/view/7821/6411>.
- Петричко М. В. Автоматична категоризація науковців за тематикою досліджень на основі профілей в Google Scholar [Електронний ресурс] / М. В. Петричко, С. Д. Штовба // ВНТУ. – 2018. – Режим доступу до ресурсу: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2018/paper/view/5427/4433>.

Публікації

- Shtovba S., Petrychko M. Jaccard index – Based Assessing the Similarity of Research Fields in Dimensions / 1st International Workshop on Digital Content & Smart Multimedia 2019, Lviv. In press.
- Shtovba S., Shtovba O., Petrychko M. Impact of the syntactic dependencies in the sentences on the quality of the identification of the toxic comments in the social networks / Scientific Works of Vinnytsia National Technical University, Vinnytsia, Ukraine, 2019.
- Shtovba S., Shtovba O., Petrychko M. Detection of Social Network Toxic Comments with Usage of Syntactic Dependencies in the Sentences / Proc. of the Second International Workshop on Computer Modeling and Intelligent Systems, Zaporizhzhia, Ukraine, April 15-19, 2019. CEUR Workshops Proceeding, Vol. 2353. – 2019. – P. 313-323.CMIS-2019.

ВИСНОВКИ

Мета роботи досягнута :

- розроблено модель представлення науковців у просторі галузей та спеціальностей наук;
- досліджено 4 системи класифікації наук та обгрунтовано вибір ANZSRC. для вирішення задачі категоризації;
- розроблено модель їх оцінювання схожості наукових спеціальностей на основі індекса Жакара та експериментально розраховано схожість для 10878 пар спеціальностей;
- покращено алгоритм категоризації науковців за спеціальностями, який враховує взаємодію як інтересів науковців, так і наукових спеціальностей;
- реалізовано програмно запропоновані моделі та алгоритми.