

Візуалізації прихованих відношень за профілями науковців в Google Scholar

Вінницький національний технічний університет

Анотація

Проведено збір та обробку даних за профілями науковців в Google Scholar. Встановлено відповідність між їхніми напрямками дослідження, та проведено візуалізацію за допомогою програми Gephi, що дозволяє побудувати карти науки за профілями науковців.

Ключові слова: візуалізація, наукова мережа, карта науки, Google Scholar.

Abstract

A data collection for scientists profiles in Google Scholar. Established correspondence between their areas of study and conducted visualization using Gephi, allowing you to build maps of science for scientists profiles.

Keywords: visualization, scientific network, map of science, Google Scholar.

Вступ

Візуальна інформація великої кількості даних краще сприймається і дозволяє швидко і ефективно зрозуміти її. Фізіологічно, сприйняття візуальної інформації є основною для людини [1]. Є численні дослідження, які підтверджують, що в 60 000 разів швидше сприймається візуальна інформація в порівнянні з текстовою. Візуалізація інформації спрямована на створення підходів до передачі абстрактної інформації в інтуїтивно зрозумілі способи.

Подібним дослідженням стали відомі своєї новаторською роботою в художній та соціальній візуалізації даних Фернанда Вієгас і Мартін Ваттенберг. Вони представляють дослідницьку групу, яка займається візуалізацією даних Google. Вони заснували області соціального аналізу даних і були творцями "Many Eyes". У своїй новій роботі за допомогою нових можливостей браузерів Chrome, Firefox та програмування у JavaScript, вони розробили інтерактивну, анімовану, дуже гарну карту США, на якій показані усі вітрові потоки країни (одразу видно де можна робити електричні вітрові станції). Їх робота була показана в музеях по всьому світу, і допомогла перетворити візуалізацію на художню практику [2].

Мета моєї роботи є розробка візуалізації прихованих відношень за профілями науковців в Google Scholar. Дана розробка надасть можливість зробити аналіз сайту Google Scholar по напрямках дослідження науковців у простій та читабельній формі для розуміння.

1. Постановка задачі та ідея дослідження

Візуалізація даних – це наочне представлення масивів різної інформації. Задача полягає в тому, щоб візуалізувати приховані відношення за профілями науковців в Google Scholar. Це дозволить оцінити напрямки дослідження науковців, та зробити статистику інтересів чим зацікавленні дослідники. Крім того, наявність даних є необхідною умовою для візуалізації. Під час візуалізації можна вирішити задачі аналізу сайту Google Scholar по профілях науковців.

З пошукової система, яка індексує повний текст наукових публікацій всіх форматів і дисциплін Google Scholar було імпортовано базу даних по профілях науковців [3]. База даних склала 831 науковців.

Для візуалізації даних повинно бути дві таблиці: nodes и edges.

Таблиця nodes повинна містити колонки:

- id – номер вершини;
- label – позначка або позначення вершини;
- count – кількість.

Для таблиці edges колонки наступні:

- source – id початкової вершини;
- target – id кінцевої вершини;
- weight – вага ребра;

– type – орієнтований, неорієнтований.

2. Приклад візуалізації

Підготовка даних

Формуємо та посилаємо на сервер запит, щоб отримати дані по науковцях, а саме їх напрямки дослідження. Зберігаємо у файл (рисунок 1).

1	fuzzy stability and control systems
2	fuzzy IT services
3	Neural Networks Control Theory Fuzzy Cryptography Optimal control
4	fuzzy
5	quality reliability maintenance fuzzy system
6	Mathematical Modelling Soft Computing and Machine Intelligence Vibration and Inverse Vibration Problems
7	Dizi uzaylar Fuzzy dereceden istatistiksel
8	Computer vision Pattern classification Machine learning Soft computing Fuzzy
9	Fuzzy ANN Control Robotics AI
10	Operations Research DEA MCDM Fuzzy Supply Chain Management
11	image processing fuzzy intuitionistic fuzzy mathematics medical image processing
12	IT Project Management Databases Fuzzy Data mining
13	Power Electronics Control Engineering Electrical Machines Fuzzy Neural networks
14	probabilita statistica fuzzy option pricing
15	control sampled-data fuzzy digital redesign LMI
16	Supplier selection multicriteria decision making fuzzy information technology knowledge management
17	AHP Fuzzy
18	Software Engineering Fuzzy Computer Graphics Softcomputing
19	Fuzzy Mathematics fuzzy sets fuzzy
20	Data Envelopment Analysis Fuzzy Aggregating preference ranking Ecological Modelling
21	Information Technology CS Petri nets Fuzzy Risk Management
22	artificial intelligence image processing segmentation fuzzy clustering
23	Fault Tolerant Control Systems Adaptive Control of Time Delay-Systems System Identification Fractional ordered systems Fuzzy

Рисунок 1 – Файл даних початкових

Обробка даних за допомогою програми, що написана на мові програмування C#: створення файлу «nodes.csv» та створення файлу «edges.csv» (рисунок 2).

1	id	label	count	1	Source	Target	Type	Weight
2	0	dereceden	1	2	85	208	Undirecte	1
3	1	aandroid	1	3	85	119	Undirecte	1
4	2	acdrives	1	4	147	38	Undirecte	1
5	3	adhoc	1	5	147	85	Undirecte	1
6	4	adaptivec	1	6	147	43	Undirecte	1
7	5	aggregatir	1	7	147	156	Undirecte	1
8	6	agile	1	8	38	85	Undirecte	1
9	7	agronomi	1	9	38	43	Undirecte	1
10	8	ahp	2	10	38	156	Undirecte	1
11	9	ai	1	11	85	43	Undirecte	1
12	10	anfis	2	12	85	156	Undirecte	1
13	11	ann	1	13	43	156	Undirecte	1
14	12	antcolony	1	14	85	85	Undirecte	8
15	13	appliedm	1	15	177	182	Undirecte	1
16	14	arduino	1	16	177	129	Undirecte	1
17	15	artificialin	9	17	177	85	Undirecte	2
18	16	artificialn	1	18	177	217	Undirecte	1
19	17	automatic	1	19	182	129	Undirecte	1
20	18	bigdata	1	20	182	85	Undirecte	1
21	19	biomedica	2	21	182	217	Undirecte	1
22	20	bss	1	22	129	85	Undirecte	1
23	21	cf	1	23	129	217	Undirecte	1

Рисунок 2 – Файл вузлів та ребер

Візуалізація за допомогою Gephi

Візуалізація даних проходить у декілька етапів:

- Імпорт даних у «лабораторію даних».
- Обробка даних (рисунок 3).

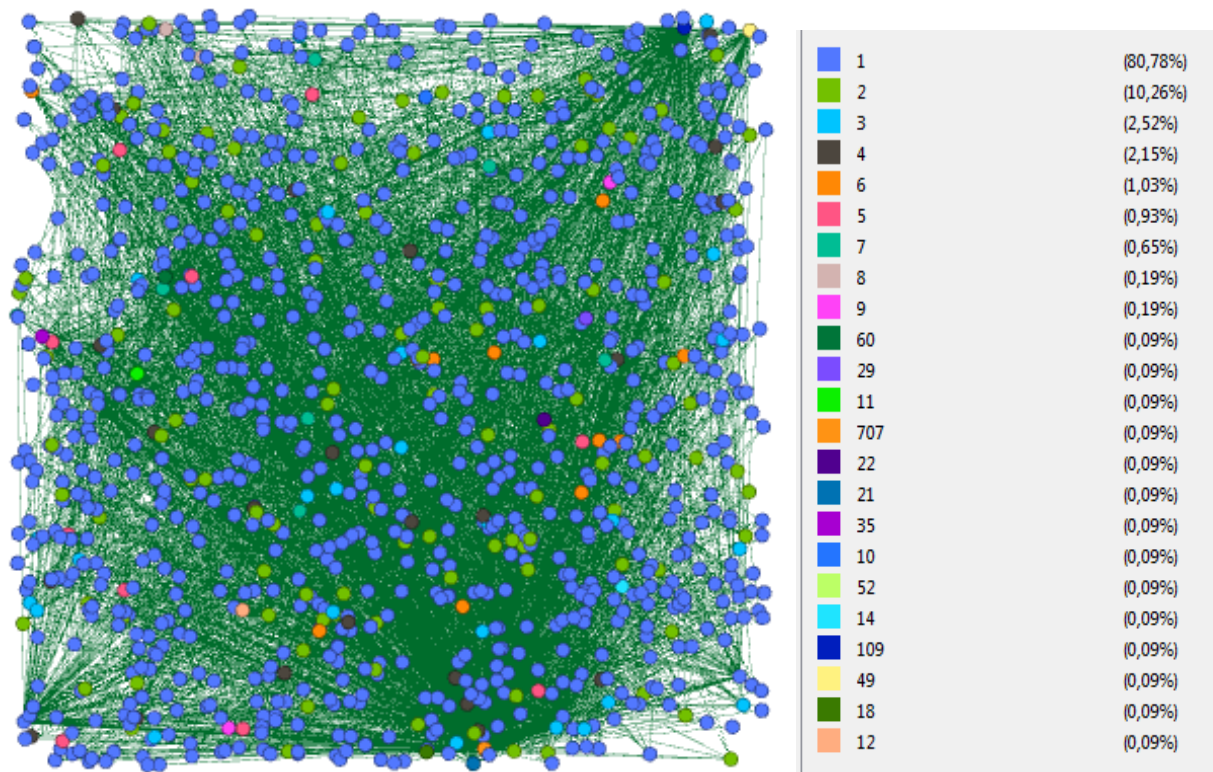


Рисунок 3 – Результат обробки даних

– Перегляд графу (рисунок 4 - 5).

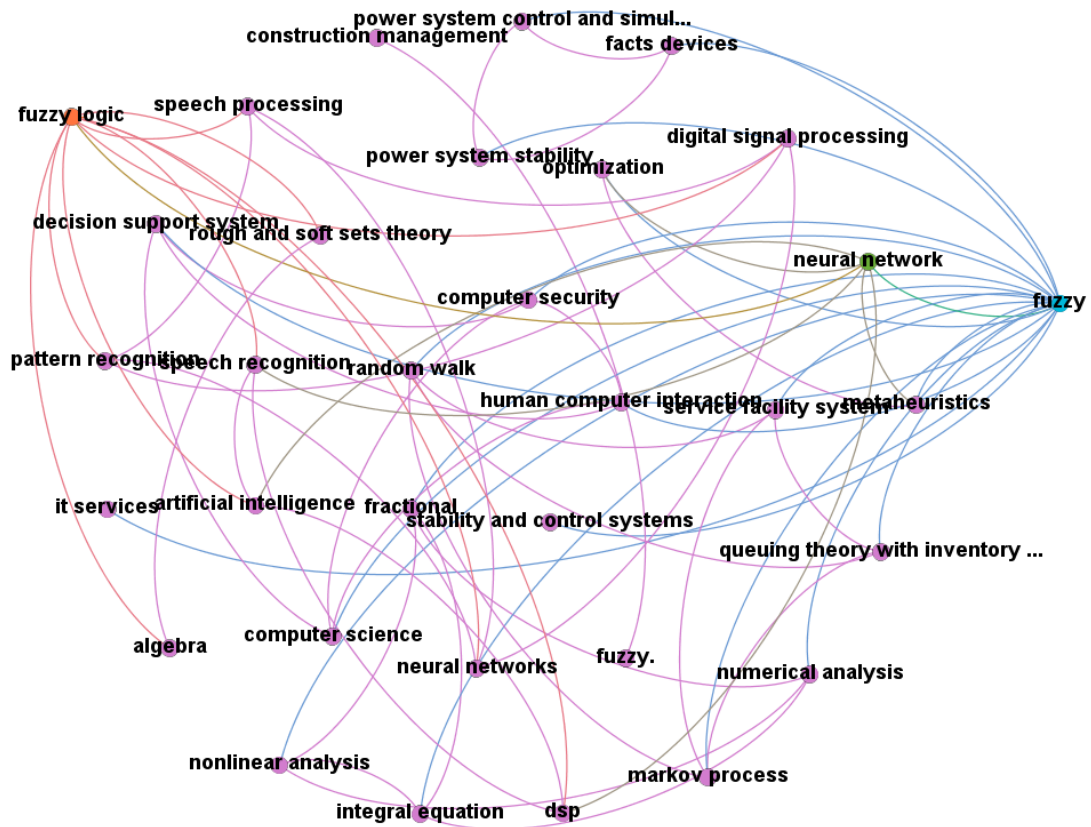


Рисунок 5 – Результат візуалізації даних

Висновок

Проведено збір та обробку даних за профілями науковців в Google Scholar. Встановлено відповідність між їхніми напрямками дослідження, та проведено візуалізацію за допомогою програми Gephi, що дозволяє побудувати карти науки за профілями науковців.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Паклин Н. Б., Орешков В. И. Визуализация данных // Бизнес-аналитика. От данных к знаниям. – 2-е изд. – СПб.: Питер, 2013. –173–210с.
2. Krum R. Cool infographics: effective communication with data visualization and design. – Indianapolis: Wiley, 2014. – 348 p.
3. Fernanda Viegas & Martin Wattenberg "Many Eyes": [Електронний ресурс] / Fernanda Viegas & Martin Wattenberg – Режим доступу до ресурсу: <http://hint.fm/wind/>.

Манглієва Тетяна Назарівна – факультет комп'ютерних систем і автоматики, 1АВ-12б, Вінницький національний технічний університет, м. Вінниця, manglievatanyha@gmail.com

Науковий керівник: **Штовба Сергій Дмитрович** – д-р техн. наук, професор, професор кафедри комп'ютерних систем управління, Вінницький національний технічний університет, м. Вінниця, shtovba@ksu.vntu.edu.ua

Tetiana Manhlieva – Faculty for Computer Systems and Automation, 1AV-12b, Vinnytsia National Technical University, Vinnytsia, manglievatanyha@gmail.com

Supervisor: **Serhiy Shtovba** – Doctor of Science, Professor, Computer Control Systems Department, Vinnytsia National Technical University, Vinnytsia, shtovba@ksu.vntu.edu.ua