

СИСТЕМАТИЗАЦІЯ ПОВНОТЕКСТОВИХ НАДХОДЖЕНЬ ДО ЕЛЕКТРОННОЇ БІБЛІОТЕКИ МЕТОДАМИ КЛАСТЕРИЗАЦІЇ

Віктор Заїка,

Національна бібліотека України імені В. І. Вернадського

Олександр Кузнєцов,

Державна науково-технічна бібліотека України

Віктор Гандзюк,

Державна науково-технічна бібліотека України

Анотація. Мета статті – запропонувати метод автоматичної систематизації (предметизації) повнотекстових надходжень, застосовуючи сучасні методи кластеризації, і на реальному прикладі реалізувати його. Отримати певну кількість кластерів для масиву файлів.

Методологія дослідження. Застосовано кількісний метод обробки даних. Для кожного файлу застосовано метод векторизації, який полягає у формуванні векторів із текстових файлів із подальшим використанням методів векторної алгебри.

Наукова новизна. Запропоновано аналітичний метод для автоматичного знаходження кластерів у масиві текстів.

Висновки. Перевагою цього методу є незалежність від людського фактору. Розглянуто та практично застосовано метод кластеризації для виділення подібних за тематикою груп вхідних документів, що може бути корисним для створення та наповнення електронної бібліотеки.

Ключові слова: електронний документ; електронна бібліотека; систематизація; векторизація; кластер, кластеризація.

ВСТУП

Сучасні вебтехнології: переклад з однієї мови на іншу, синхронний переклад у реальному часі з технологіями перетворення звуку в текст і тексту у звук, пошук помилок у словах та побудові речення, транслітерація, статистична обробка та аналіз даних вільно доступні та необхідні для людини науки. Використання різноманітних методів кластеризації зробили пошук в інтернеті більш гнучким і точним порівняно з пошуком за ключовими словами.

В останні роки нові надходження до бібліотек часто мають цифровий вигляд — текстових файлів або тестових шарів більш складних документів, наприклад збережених у pdf-форматі. Тому вебтехнології можуть суттєво спростити вирішення різних завдань сучасної бібліотеки.

Уже є чимало програмних засобів аналізу текстів і написано багато статей із застосуванням методик такого аналізу. Ідеться здебільшого про контент соціальних мереж, створений мовою програмування python. Згенерований у такий спосіб матеріал використовують зазвичай так звані корпуси сумнівного походження, мабуть, із тих самих соціальних мереж, що зводить нанівець застосування створеної інформації для бібліотечних потреб.

Новітні технології обробки текстів ще далекі від систематизації цих текстів на рівні бібліотек. Колеги бібліотекарів із кіберпростору, які теж працюють з інформацією, технологічно змогли повторити тільки абетковий каталог. У площині ж систематичного каталогу та його електронних аналогів фахівці електронних технологій ще жодних результатів не досягли. І навіть пошукова система

Google працює швидко, але неточно, позиціонує можливість вказувати під пошуковим запитом кількість наносекунд, витрачених на виконання запиту, як свою ключову перевагу. Ось чому деякі вчені (істинні вчені, а не генератори статей, які з легкістю проходять сучасні системи антиплагіату) віддають перевагу роботі в бібліотеці й шукають інформацію в картотеках та бібліотечному електронному каталозі. Адже для таких користувачів пріоритетною є якість (адекватність / точність) пошуку. Однак еволюція інформації пішла в бік швидкості. Саме ця тенденція стимулює роботу бібліотекарів у напрямі перенесення бібліотечних технологій у комп'ютерний простір.

Комп'ютерні фахівці часто не мають уявлення про значний багаторічний досвід і досягнення бібліотекарів у сфері обробки, аналізу та пошуку інформації, у той час як бібліотекарі й видавці можуть не знати про результати та напрями досліджень, що проводяться розробниками інформаційних технологій у цій самій сфері [3].

Розглянемо питання систематизації повнотекстових надходжень: предметні рубрики, Універсальну десяткову класифікацію (УДК) та можливість автоматизації процесів. З'ясуємо, скільки предметних рубрик і підрубрик приходить на певну кількість повнотекстових документів.

Як правило, систематизатор ознайомлюється тільки з назвою та анотацією, оскільки прочитати весь документ повністю не вистачає часу та фізичних можливостей. Перелічені фактори є основними причинами хибного присвоєння індексу УДК.

Системи аналізу тексту (контенту) позбавлені перелічених недоліків. Тобто комп'ютерна програма з легкістю за лічені секунди чи навіть долі секунди може неупереджено й повністю обробити текстовий файл будь-якого розміру та виокремити характеристики, які лежать в основі пошуку.

Отже, постає питання про очевидну доцільність поєднання переваг людини та комп'ютера — озброєння бібліотекаря новими технологіями, без яких йому в подальшій роботі не обійтися.

Для обробки великих масивів цифрового тексту застосовуються багато програм і методик: вони допомагають упоратися з масштабітністю даних, але не завжди здатні визначити тематику отриманих для опрацювання текстів.

Для визначення тематики цифрового тексту в цій роботі застосовується метод векторизації текстових файлів [1].

На рис. 1 зображено чотири етапи процесу векторизації: текстові файли, масиви слів, лексеми ttt - t, частотні масиви.

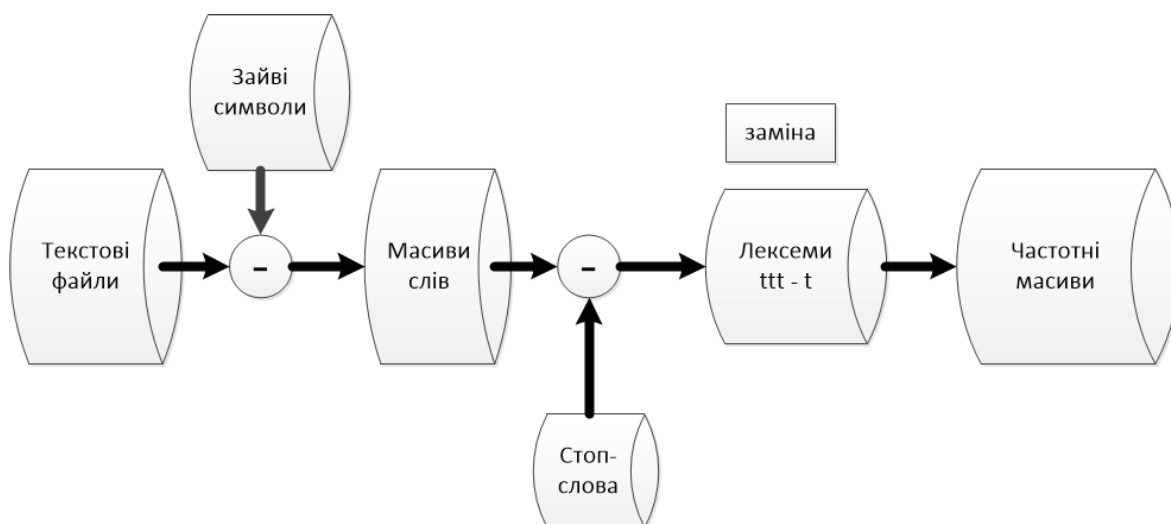


Рис. 1. Схема векторизації текстових файлів

файли надходять до програми й відбувається вилучення зайвих символів, які будуть заважати під час обробки тексту. На другому етапі формується масив слів, узятих із тексту, а потім стоп-слова вилучаються з масиву. Стоп-слова — це окремі слова чи фрази, позбавлені конкретного сенсу, які лише перервантажують текст надлишковими лінгвістичними конструкціями. Процес знаходження українських стоп-слів повною мірою залежить від людського фактору. На третьому етапі в утвореному масиві слів відбувається заміна слів на лексеми. Лексема являє собою основну одиницю лексичного значення, яка не залежить від можливих флективних закінчень або варіацій у формах слова. У словнику основні слова є лексемами. Із погляду формального морфологічного аналізу в лінгвістиці лексема абстрактна й приблизно відповідає різноманітним формам одного й того самого слова. Синоніми та словоформи замінюються на одну обрану лексему як головну.

Інтерфейс модуля «Лексеми» (рис. 2) програми «Аналіз текстів» дає більше уявлення про згаданий вище процес заміни слів на лексеми: усі одиниці поля «Лексеми» замінюються на слово, вказане в лівому полі.

На останньому етапі для кожного файлу створюються частотні словники, які зберігаються в таблицях, упорядкованих за частотою використання слова.

Кластеризація — це статистичний аналіз, який розбиває об’єктні вибірки на декілька підмножин — кластерів, — які між собою не перетинаються [3]. З об’єктів, які схожі по своїй суті, повинні складатися практично всі кластери, а якщо кластери різні, то їхні об’єкти також повинні бути різними. У нашому випадку повнотекстова вибірка — 35 файлів, які треба розподіляти на групи (кластери) за тематичними напрямками. Оскільки програма працює тільки з текстовими шарами документів, наведемо їх список (Рис. 3).

МЕТОДОЛОГІЯ ДОСЛІДЖЕННЯ

Можна стверджувати, що завдання систематизації являє собою завдання кластеризації

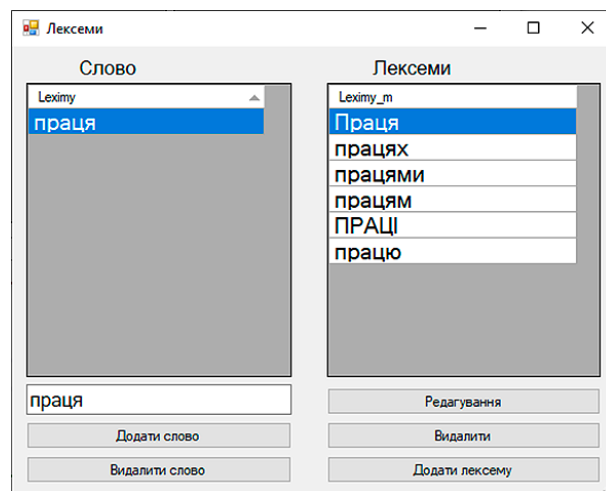


Рис. 2. Модуль «Лексеми»

Имя	Дата изменения	Тип	Размер
001.txt	07.09.2023 12:11	Текстовый докум...	856 КБ
002.txt	07.09.2023 16:04	Текстовый докум...	1 237 КБ
003.txt	15.06.2023 12:25	Текстовый докум...	576 КБ
004.txt	15.06.2023 12:26	Текстовый докум...	576 КБ
005.txt	15.06.2023 12:27	Текстовый докум...	905 КБ
006.txt	07.09.2023 15:40	Текстовый докум...	496 КБ
007.txt	07.09.2023 16:12	Текстовый докум...	974 КБ
008.txt	07.09.2023 16:13	Текстовый докум...	1 651 КБ
009.txt	15.06.2023 12:35	Текстовый докум...	1 630 КБ
010.txt	15.06.2023 12:37	Текстовый докум...	1 632 КБ
011.txt	15.06.2023 12:39	Текстовый докум...	1 065 КБ
012.txt	15.06.2023 12:40	Текстовый докум...	581 КБ
013.txt	21.06.2023 13:30	Текстовый докум...	627 КБ
014.txt	21.06.2023 13:31	Текстовый докум...	556 КБ
015.txt	21.06.2023 13:32	Текстовый докум...	478 КБ
016.txt	21.06.2023 13:33	Текстовый докум...	490 КБ
017.txt	21.06.2023 13:33	Текстовый докум...	151 КБ
018.txt	21.06.2023 13:34	Текстовый докум...	811 КБ
019.txt	21.06.2023 13:35	Текстовый докум...	1 748 КБ
020.txt	21.06.2023 13:36	Текстовый докум...	2 436 КБ
021.txt	21.06.2023 13:36	Текстовый докум...	213 КБ
022.txt	21.06.2023 13:37	Текстовый докум...	123 КБ
023.txt	21.06.2023 13:38	Текстовый докум...	749 КБ
024.txt	21.06.2023 13:39	Текстовый докум...	318 КБ
025.txt	21.06.2023 13:39	Текстовый докум...	7 КБ
026.txt	21.06.2023 13:40	Текстовый докум...	34 КБ
027.txt	21.06.2023 13:41	Текстовый докум...	27 КБ
028.txt	21.06.2023 13:41	Текстовый докум...	32 КБ
029.txt	21.06.2023 13:42	Текстовый докум...	35 КБ
030.txt	21.06.2023 13:42	Текстовый докум...	43 КБ
031.txt	21.06.2023 13:43	Текстовый докум...	16 КБ
032.txt	21.06.2023 13:43	Текстовый докум...	20 КБ
033.txt	21.06.2023 13:44	Текстовый докум...	32 КБ
034.txt	21.06.2023 13:48	Текстовый докум...	2 730 КБ
035.txt	22.06.2023 16:55	Текстовый докум...	40 КБ

Рис. 3. Список файлів — текстові шари документів

зації, тому що розбиття об'єктів між собою відбувається завдяки схожості їхніх векторних ознак. Це завдання є загальним, тому в кожному випадку використовують різні методи кластеризації. Утворення кластерів відбувається за рахунок того, що між векторами є відстань, або завдяки щільним ділянкам в інтервалах, або в просторі даних, або точних розподілах статистичного типу. Чіткий набір даних і ціль використання результатів – це два фактори, від яких залежить усе вище-описане. Метод кластерного аналізу не є автоматичним — він являє собою ітераційну процедуру: метод, за яким опрацьовуються дані, і характеристики моделі змінюються до того моменту, поки не вдасться отримати результат згідно із заданими властивостями.

Розв'язання є неоднозначним із декількох причин. По-перше, у кластеризації немає показника, за яким можна було б охарактеризувати її якість. Є група ефективних алгоритмів і характеристик, яким, проте, бракує чіткого критерію. Незважаючи на це, у будь-якому випадку через ці алгоритми здійснюється доволі якісна кластеризація з побудови. Кожен із них може привести до отримання різноманітних результатів.

Міра подібності числових векторів (або косинусна міра), які ще називаються скалярним добутком, ділиться на добуток модулів, що відносяться до них. По-іншому, це означає, що два задані вектори A і B косинус-подібності, $\cos(\theta)$, можуть представлятися через довжину та скалярний добуток.

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (1)$$

У цій формулі A_i та B_i — координатами, які відносяться до вектора A та B . Подібність косинусна залежить від кута між векторами, а не лише від величини векторів. Значенням косинуса кута може стати число від -1 до 1 . При цьому значенням косинусної подібності двох векторів, які розташовані пропорційно, є 1 — обох перпендикулярних векторів, а 0 і -1 — обох векторів, які протилежні один від одного. Значення, які належать координатам

векторів, не будуть мінусовими в плані аналізу даних текстового типу, бо стан подібності косинуса є обмеженим замкненням інтервалу від 0 до 1 [2].

Тематичний напрям аналізованих файлів заздалегідь не був відомий. Коефіцієнт тематичного напрямку (КТН) кожного текстового файлу визначається відносно свого корпусу, текстовий вектор якого формується як сума векторів з усіх файлів, крім аналізованого: у разі збігу слів слово замінюється, а частоти додаються, а в разі не збігу в сумарний вектор додається і слово, і його частота. Тобто у формулу (1) підставляються вектор відповідного файлу та його корпус. Отримані значення впорядковуються за значенням КТН. На рис. 4 можна легко побачити кластери вхідних 35 текстів — майже горизонтальні полицки, обведені овалами. Упорядкований одновимірний масив чисел, більших за нуль, значно спрощує пошук кластерів: декартова відстань спрощується до простого віднімання координат. А впорядкування чисел спрощує обхід усіх координат — коефіцієнтів КТН: можна справа наліво або навпаки.

З математичних міркувань умовою знаходження файлів у кластері є:

$$|K_i - K_{i+1}| \leq \varepsilon, \quad (2)$$

де K_i , K_{i+1} — коефіцієнти КТН сусідніх файлів, ε — допустима величина зміни функції, КТН — критерій кластеру, значення від якого залежать кількість кластерів і точність визначення тематики (КТН).

Із Рис. 5 видно, що похідна функції КТН у точці чисельно дорівнює різниці K_{i+1} та K_i , тому що відстань між ними по осі Ox дорівнює одиниці. Щоб оцінити певне значення ε з формули (1) треба обрати мінімальне значення ε при якому не виконується умова знаходження в одному кластері, тобто перехід у наступний, яка дорівнює $|K_i - K_{i+1}|$. Для цього побудуємо графік функції похідної графіка КТН (Рис. 5): вісь Ox — номери файлів, а Oy — абсолютна різниця $K_i - K_{i+1}$.

Розглянемо похідну функції КТН — червоний колір (Рис. 6). Вона має критичні точки, серед яких чотири основні, їх значення змінюються більш як на 10 – 15% — це і є межі

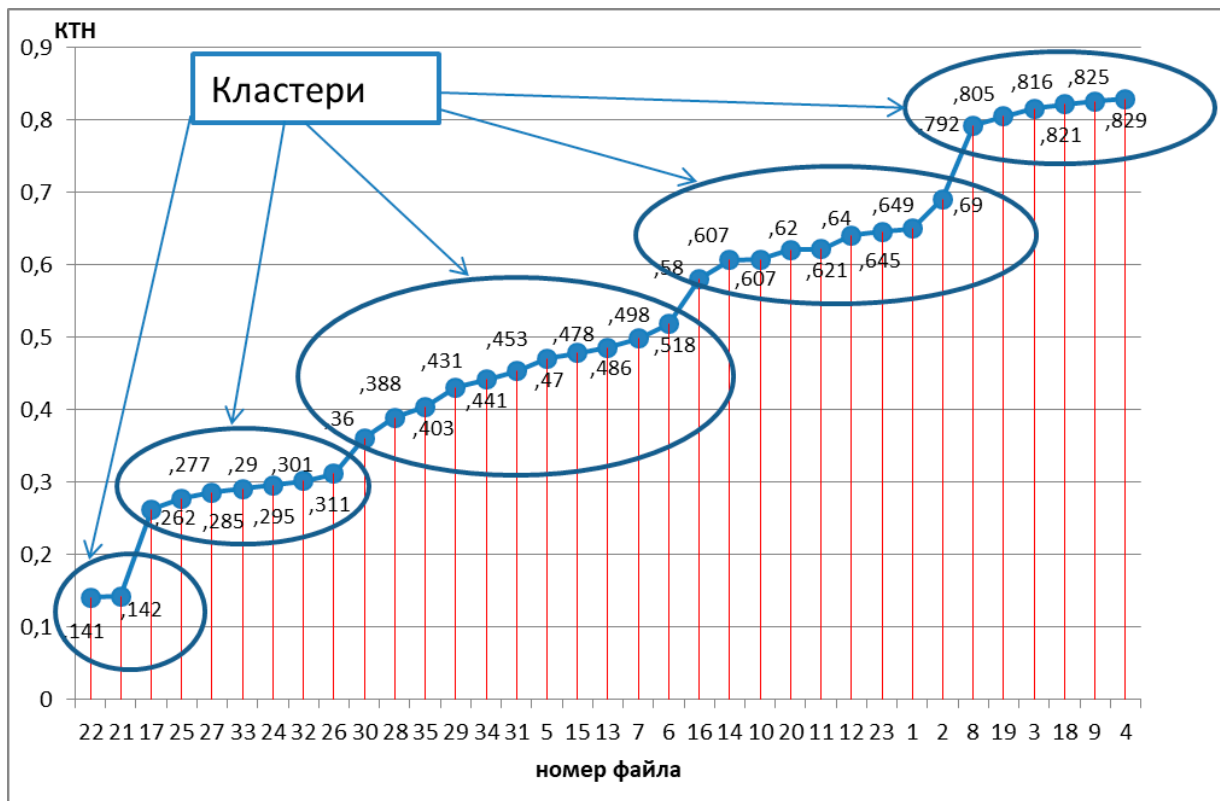


Рис. 4. Розподіл КТН розглянутих файлів електронній бібліотеці ДНТБ України

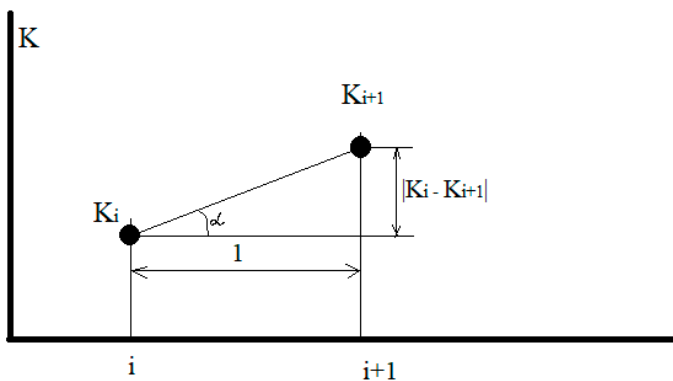


Рис. 5. Координати двох сусідніх точок

кластерів. Для критерію кластера для формули 1 очевидно обрати із чотирьох цих точок точку А, яка має мінімальне значення. Цей перехід до наступного кластеру відбувся на файлах 26 та 30. Файл 26 зі значенням КТН 0,311 (Рис. 4), та наступний файл 30 зі значенням КТН 0,36. Тут значення ε дорівнює $0,36 - 0,311 = 0,049 \approx 0,05$. Отже для нашого випадку оптимальна $\varepsilon = 0,05$ і її можна знайти не тільки графічно, но і аналітично.

Алгоритм знаходження кластерів після знаходження критерію ε який: обійти впорядкований масив значень КТН файлів за

зростанням їх коефіцієнтів. Якщо виконується умова формули 1, то ці файли належать до одного й того самого кластеру. Таким чином, вхідний масив поділяємо на підмножини файлів – кластери. І очевидно, що в межах методу в цілому відбувся перехід від графоаналітичного методу до аналітичного.

Виявлення кластерів можна вважати достовірним, тому що втручання людини було мінімальним. Для перевірки роботи програми було взято два файли — 9 і 4 (Рис. 4 праворуч) два файли одного й того самого кластеру (див. табл. 1). Кожний із цих файлів має текст зі своєю тематикою, і програма має визначити цю тематику.

У файлі №9: Колективна монографія присвячена науковому обґрунтуванню шляхів вирішення проблем соціально-економічної дивергенції України та держав Європейського Союзу, впровадження підходів державної політики, спрямованої на стимулювання їх конвергенційної взаємодії. Показано ключові виклики євроінтеграційних процесів України, у т.ч. обумовлені турбулентністю сучасного глобального середовища, набутою під впли-

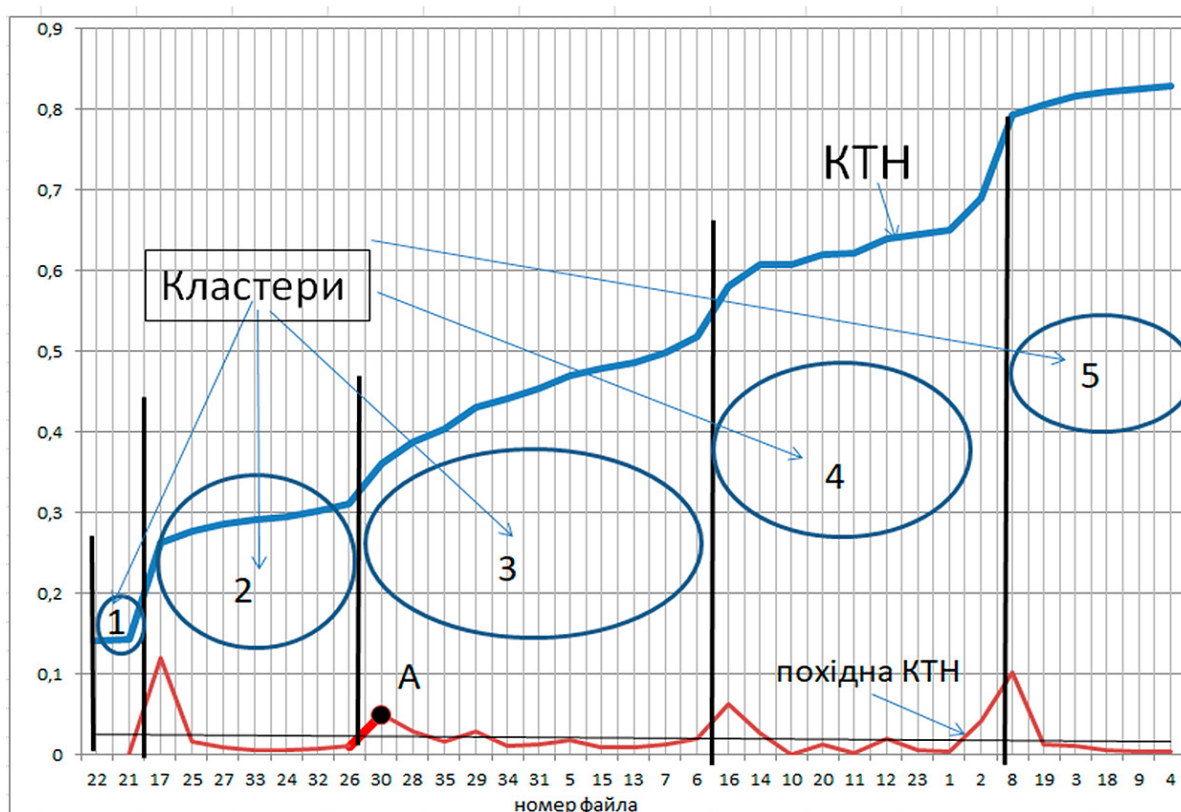


Рис. 6. Похідна функції КТН

вом різних чинників макроекономічною вразливістю нашої держави, а також посиленням дезінтеграційних тенденцій у розвитку Євросоюзу через низку невирішених інституційних та ін. проблем. Аргументовано необхідність імплементації Угоди про асоціацію між Україною та ЄС крізь призму забезпечення їх соціально-економічної конвергенції. Розкрито суть і запропоновано підходи до вирішення системних проблем конвергенції України та країн Євросоюзу за секторами (сферами) економіки. Обґрунтовано концептуальні засади та моделі протидії дивергенційним тен-

денціям у процесі інтеграції України до об'єднаної Європи. Монографія буде корисною фахівцям у сфері міжнародної економіки, макроекономічного, фінансового та інституційного розвитку, державного управління, а також викладачам і здобувачам вищої освіти за економічними спеціальностями.[4]

Отже, тематикою цього тексту є економіка — УДК 339.92.

У файлі №4: У збірнику розкривається актуальність теоретичного дослідження процесів світової економіки (швидке зростання вільного потоку товарів і послуг, капіталу,

Таблиця 1

Файли, які потрапили до одного кластеру

Ім'я файлу (текстовий шар)	Назва праці	УДК	Розмір тексту, кб
004.txt	Розвиток в Україні досліджень в галузі міжнародної економіки: Збірник наукових тез доповідей	339.9:167(477)	576
009.txt	Оцінка інтеграції України до європейського економічного простору	339.92(477+945.4)-045.74	1630

Таблиця 2

Список слів із кількістю повторень, проаналізованих із файлів 9 та 4

Файл №9		Файл №4	
країна	903	країна	220
розвиток	586	економічна	218
ринок	462	економіка	217
економічна	451	розвиток	210
економіка	371	система	101
система	270	наука	101
праця	249	перевезення	100
сфера	242	міжнародних	91
зайнятість	194	ринок	80
зростання	192	відносини	80
товар	188	досліджень	77
політики	188	сфера	75
Угоди	188	політики	66
торгівлі	154	забезпечення	57
виробництва	154	управління	55
асоціацію	143	світ	54
заход	142	заход	51
Україною	142	праця	50
діяльності	141	наукових	50
послуг	139	дослідження	50

робочої сили і фінансів між країнами). Висвітлено бачення міжнародних економічних проблем та відмінності у їх дослідженні науковцями академічних інститутів НАН України та освітянами у вищих навчальних закладах МОН України на відповідних кафедрах. Проаналізовано наслідки глобалізації – негативні та позитивні з урахуванням проблем та завдань сьогодення.[5]

Тож тематика цього тексту — теж економіка — УДК 339.9.

Отже, за допомогою програми було проаналізовано два текстових файли та визначено, що їх тематикою є економіка і що тематика цих файлів збігається.

У табл. 2. виявлені слова та кількість їх повторень у файлах. За цими файлами було встановлено, що спільним тематичним напрямом є економіка. Тому навіть якби систематизатор, тобто людина, прочитала ці два

файли, то так само підтвердила б належність обох файлів до спільної тематики, що підтверджує правильність запропонованого методу.

ВИСНОВКИ

Запропонували аналітичний метод кластеризації (предметизації), який не залежить від людського фактору. На конкретному прикладі виявили п'ять кластерів для 35 досліджуваних файлів. Ця інформація буде корисна для бібліотекарів-систематизаторів, задіяних у процесі визначення предметних рубрик.

Частотні масиви можуть бути використані для подальшої роботи із систематизації повнотекстових надходжень.

Поповнено словник стоп-слів та словник синонімічних лексем.

Доцільно було б створювати тематичні корпуси для лінгвістичних досліджень, використовуючи повнотекстові ресурси бібліотек.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Визначення індексів УДК нових надходжень [Текст] / О. Ю. Кузнецов, В. Заїка; Бібліотечний вісник // Бібліотечний вісник. — 2023. — No: 3. — Ст. 1. — С. 3–16
2. Лобузін І. Цифрові бібліотечні проекти: технологічні рішення та управління життєвим циклом колекцій : монографія / Іван Лобузін ; відп. ред. В. А. Широков ; НАН України, Нац. б-ка України ім. В. І. Вернадського. — Київ, 2016. — 216 с.
3. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008, URL:<https://nlp.stanford.edu/IR-book/pdf/06vect.pdf>
4. Оцінка інтеграції України до європейського економічного простору : колективна монографія / за заг. ред. д.е.н., проф. О.О. Борзенко ; НАН України, ДУ «Інститут економіки та прогнозування НАН України». — Електрон. дані. — К., 2021. — 493 с. : табл., рис. — Режим доступу : <http://ief.org.ua/wp-content/uploads/2022/06/Оцінка-інтеграції-України- до- ЕЕП-2.pdf>
5. Розвиток в Україні досліджень в галузі міжнародної економіки : збірник наукових тез доповідей дискусійної платформи 28 вересня 2021 р. / НАН України, ДУ «Ін-т екон. та прогнозув. НАН України». — Електрон. дані. — К., 2021. — 127 с. — Режим доступу : <http://ief.org.ua/docs/scc/20.pdf>

REFERENCES

1. Vyznachennia indeksiv UDK novykh nadkhodzen [Tekst] / O. Yu. Kuznietsov, V. Zaika; *Bibliotechnyi visnyk* // *Bibliotechnyi visnyk*. — 2023. — No 3. — St. 1. — S. 3–16 [In Ukr.]
2. Lobuzin I. Tsyfrovi bibliotechni proekty: tekhnolohichni rishennia ta upravlinnia zhyttievym tsyklom kolektsii : monohrafiia / Ivan Lobuzin ; vidp. red. V. A. Shyrokov ; NAN Ukrainy, Nats. b-ka Ukrainy im. V. I. Vernadskoho. — Kyiv, 2016. — 216 s. [In Ukr.]
3. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008, URL:<https://nlp.stanford.edu/IR-book/pdf/06vect.pdf>
4. Otsinka intehratsii Ukrainy do yevropeiskoho ekonomichnoho prostoru : kolektyvna monohrafiia / za zah. red. d.e.n., prof. O.O. Borzenko ; NAN Ukrainy, DU «Instytut ekonomiky ta prohnozuvannia NAN Ukrainy». — Elektron. dani. — K., 2021. — 493 s. : tabl., rys. — Rezhym dostupu : <http://ief.org.ua/wp-content/uploads/2022/06/Otsinka-intehratsii-Ukrainy-do-EEP-2.pdf> [In Ukr.]
5. Rozvytok v Ukraini doslidzhen v haluzi mizhnarodnoi ekonomiky : zbirnyk naukovykh tez dopovidei dyskusiinoi platformy 28 veresnia 2021 r. / NAN Ukrainy, DU «In-tekon. ta prohnozuv. NAN Ukrainy». — Elektron. dani. — K., 2021. — 127 s. — Rezhym dostupu : <http://ief.org.ua/docs/scc/20.pdf> [In Ukr.]

CLUSTERIZATION OF RECEIPTS TO THE ELECTRONIC LIBRARY

Abstract. *The purpose of the article is to propose a method for automatic systematization (subjectification) of full-text findings and daily clustering methods and implement them in a real application: identify clusters from (35) input texts that have reached the DNTB, and the very groups of texts similar to the place, in order to implement automatic systematization.*

Methodology. *A simple method of data processing has been established. For the skin file, the vectorization method is used, which is based on the formation of vectors from text files, with further use of vector algebra methods.*

Scientific innovation. *An analytical method for finding clusters (subject headings) for a library catalog in the texts with minimal systematizer intervention is proposed.*

Conclusions. *The advantage of this method is its independence from the human factor. The clustering method is examined and practically developed to identify groups of input documents similar to the subject, which can be useful in the creation of an external electronic library.*

Keywords: *electronic document; electronic library; systematization; vectorization; cluster; clusterization.*

ІНФОРМАЦІЯ ПРО АВТОРІВ

Віктор Заїка — кандидат фізико-математичних наук, провідний інженер відділу бібліометрії і наукометрії, Національної бібліотеки України імені В. І. Вернадського, Україна, м. Київ, 03039, просп. Голосіївський, 3; +0665501395; e-mail: victor.zayika@gmail.com; ORCID: 0009-0003-6582-6524

Олександр Кузнєцов — завідувач відділу наукового комплектування та опрацювання бібліотечних фондів Державної науково-технічної бібліотеки України, Україна, м. Київ, 03150, вул. Антоновича 180; +0964615376; e-mail: nkof@dntb.gov.ua; ORCID: 0000-0002-9902-1295

Віктор Гандзюк — молодший науковий співробітник відділу наукових досліджень та цифровізації Державної науково-технічної бібліотеки України, Україна, м. Київ, 03150, вул. Антоновича 180; +0663062957; e-mail: gandzyuk.v@dntb.gov.ua; ORCID: 0009-0004-3164-0622

INFORMATION ABOUT THE AUTHORS

Victor Zayika — candidate of physical and mathematical sciences, Engineer of the department of bibliometrics and scientometrics, of V. I. Vernadskyi National Library of Ukraine, 3, Holiivskyi Avenue, Kyiv, Ukraine, 03039; +0665501395; e-mail: victor.zayika@gmail.com; ORCID: 0009-0003-6582-6524

Oleksandr Kuznetsov — head of department of Scientific Document Processing and Cataloging of the The State Scientific and Technical Library of Ukraine, 180, Antonovycha str., Kyiv, Ukraine, 03150; +0964615376; e-mail: nkof@dntb.gov.ua; ORCID: 0000-0002-9902-1295

Viktor Gandzyuk — junior researcher of the Department of Scientific Research and Digitization of the The State Scientific and Technical Library of Ukraine, 180, Antonovycha str., Kyiv, Ukraine, 03150; +0663062957; e-mail: gandzyuk.v@dntb.gov.ua; ORCID: 0009-0004-3164-0622