

**Міністерство освіти і науки України
Одеський національний технологічний університет
Інститут комп'ютерних систем і технологій
"Індустрія 4.0" ім.П.Н.Платонова**

**«ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ І
АВТОМАТИЗАЦІЯ – 2022»**

***МАТЕРІАЛИ
XV МІЖНАРОДНОЇ НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ***



20 - 21 ЖОВТНЯ 2022 р.

м.ОДЕСА

**MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
ODESSA NATIONAL UNIVERSITY OF TECHNOLOGY
INSTITUTE OF COMPUTER SYSTEMS AND TECHNOLOGIES
"INDUSTRY 4.0" NAMED AFTER P.N. ПЛАТОНОВА**

**«INFORMATION TECHNOLOGIES AND
AUTOMATION– 2022»**

***PROCEEDINGS
OF THE XV INTERNATIONAL SCIENTIFIC AND PRACTICAL
CONFERENCE***



OCTOBER 20 - 21, 2022

ODESSA

Організаційний комітет конференції
Organizational committee of the conference

Голова
Supervisor

Єгоров Б.В., проф. (Одеса)

Заступники голови
Deputy Chairmen

Поварова Н.М., доц. (Одеса, Україна)
Хобін В.А., проф. (Одеса, Україна)
Котлик С.В., доц. (Одеса, Україна)

Члени комітету
Committee members

Panagiotis Tzionas prof. (Thessaloniki, Greece)
Qiang Huang, prof. (Los Angeles C.A., USA)
Yangmin Li, prof (Macao, China)
Артеменко С.В., проф., (Одеса, Україна)
Романюк О.Н., проф. (Вінниця, Україна)
Грабко В.В., проф. (Вінниця, Україна)
Єгоров В.Б., д.т.н. (Одеса, Україна)
Жученко А.І., проф. (Київ, Україна)
Ладанюк А.П., проф. (Київ, Україна)
Лисенко В.Ф., проф. (Київ, Україна)
Любчик Л.М., проф. (Харків, Україна)
Палов І., проф. (Русе, Болгарія)
Плотніков В.М., проф. (Одеса, Україна)
Стовкова В.Д., доц. (Тракия, Болгарія)
Суслов В., доц. (Кошалін, Польща)
Артем'єв П., проф. (Ольштин, Польща)
Судацевські В., доц. (Кишинів, Молдова)
Аманжолова С., доц. (Алмати, Казахстан)

УДК 004.01/08

Інформаційні технології і автоматизація – 2022 / Матеріали XV міжнародної науково-практичної конференції. Одеса, 20-21 жовтня 2022 р. - Одеса, Видавництво ОНТУ, 2022 р. – 246 с.

Збірник включає матеріали доповідей учасників конференції, які об'єднані за тематичними напрямками конференції.

Збірник буде корисним як для фахівців і працівників фірм, зайнятих в області ІТ, так і для викладачів, магістрів і студентів вищих навчальних закладів, які навчаються за напрямами і спеціальностями програмного забезпечення, обчислювальної техніки і автоматизованих систем, прикладної математики та обробки інформації, буде корисним професіоналам з комп'ютерного моделювання та розробки комп'ютерних ігор.

Результати досліджень у збірнику представляють собою своєрідний зріз сучасного стану справ в перерахованих галузях знань, який може допомогти як фахівцям, так і студентам університетів скласти загальну картину розвитку інформаційних технологій та пов'язаних з ними питань.

Наукові праці згруповані за напрямками роботи конференції та наведені в алфавітному порядку прізвищ авторів.

Матеріали (тези доповідей) друкуються в авторській редакції. Відповідальність за якість та зміст публікацій несе автор.

Рекомендовано для публікації Вченою Радою навчально-наукового інституту комп'ютерних систем і технологій «Індустрія 4.0» ім. П.М. Платонова ОНТУ від 27.10.2022 р., протокол № 2.

Матеріали подано українською та англійською мовами.
Редактор збірника Котлик С.В.

UDC 004.01/08

Information Technologies and Automation - 2022 / Proceedings of the XIV International Scientific and Practical Conference. Odessa, October 20-21, 2022. - Odessa, ONTU Publishing House, 2022 – 246 p.

The collection includes materials of reports of conference participants, which are united by thematic areas of the conference.

The collection will be useful for professionals and employees of companies engaged in the field of IT, as well as for teachers, masters and students of higher education institutions studying in the areas and specialties of computer software and automated systems, applied mathematics and information processing, will be useful to professionals on computer modeling and development of computer games.

The results of research in the collection are a kind of slice of the current state of affairs in these areas of knowledge, which can help both professionals and university students to get a general picture of the development of information technology and related issues.

Scientific papers are grouped by areas of the conference and are listed in alphabetical order of the authors.

Materials (abstracts) are published in the author's edition. The author is responsible for the quality and content of publications.

Recommended for publication by the Academic Council of the Educational and Scientific Institute of Computer Systems and Technologies "Industry 4.0" them. P.M. Platonov from 27.10.2022, protocol № 2.

Materials are submitted in Ukrainian and English.
Editor of the collection Sergii Kotlyk.

ЗМІСТ CONTENT

Список організацій, представники яких взяли участь у роботі конференції	12
Розділ 1. Математичне і комп'ютерне моделювання складних процесів	14
Derevianko O.I. Model of the formation of the microstructure of nanocoatings. (Oles Honchar Dnipro National University, Ukraine)	14
Акішев О.О., Арсірій О.О. Методика частотного аналізу тексту за допомогою алгоритма count-min sketch. (Національний університет «Одеська Політехніка», Україна)	17
Вербіцький В.В., Крачилова В.Д., Жарка М. С. Моделювання перенесення забруднюючих речовин у пористих середовищах. (Одеський національний університет імені І.І. Мечникова, Україна)	20
Гайдук К. С. Розробка мови опису правил онтології ТНОТН. (Національний аерокосмічний університет ім. М. Є. Жуковського "ХАІ", Україна)	21
Демент'єв А. М., Левикін В. М. Розробка моделі розрахунку прибутку підприємства. (Харківський національний університет радіоелектроніки, Україна)	24
Завальнюк Є. К., Романюк О. Н., Романюк О.В., Денисюк А.В., Котлик С.В. Аналіз рендерів для САПР. (Вінницький національний технічний університет, Одеський національний технологічний університет, Україна)	25
Каштан С.С. Математичне моделювання ідеальних та квазіідеальних полів при наявності джерела поперечних збурень. (Відокремлений структурний підрозділ «Рівненський технічний фаховий коледж Національного університету водного господарства та природокористування», Україна)	27
Козубенко М. В., Мельник О.В., Романюк О. Н., Котлик С.В. Використання гексогонального растру в картографії. (Вінницький національний технічний університет, Одеський національний технологічний університет, Україна)	30
Косолап А.І. Ефективне розв'язування мультимодальних оптимізаційних задач. (Український державний хіміко-технологічний університет, Україна)	33
Котлик С.В., Соколова О.П., Корнієнко Ю.К. Застосування математичних моделей та програмного забезпечення для проектування нових харчових продуктів (Одеський національний технологічний університет, Україна)	36
Котлов Д.Є., Свинчук О.В. Застосування методів спектрального аналізу в гідроакустиці. (Національний технічний університет України, «Київський політехнічний інститут імені Ігоря Сікорського», Україна)	40
Ракитянська Г.Б. Розробка автоматизованої системи управління ресурсами з використанням технології ML.NET. (Вінницький національний технічний університет, Україна)	42
Сохацький А.В. Математичне моделювання - засіб розробки новітніх транспортних технологій. (Інститут транспортних систем та технологій НАН України)	45
Тюріна Є. О., Ярошук Л. Д. Інформаційне забезпечення імітаційного моделювання адсорбційного очищення оливо і мастил. (Національний технічний університет України, «Київський політехнічний інститут імені Ігоря Сікорського», Україна)	48
Розділ 2. Управління, обробка та захист інформації	51
Журавська І. М., Обухова К. О. Інтелектуальна власність на вебсайтах. (Чорноморський національний університет імені Петра Могили, Україна)	51
Зінченко С.М., Товстокорий О.М., Маменко П.П., Кириченко К.В., Матейчук В.М. Використання полюсу повороту для маневрування з поздовжньою швидкістю. (Херсонська державна морська академія, Україна)	54

Список
 організацій, представники яких взяли участь у роботі конференції
 List
 organizations whose representatives took part in the conference

Masaryk University	Czech Republic
Abylkas Saginov Karaganda Technical University Kazakhstan	Kazakhstan
New Bulgarian University	Bulgaria
Taras Shevchenko National University of Kyiv	Ukraine
Turan University	Kazakhstan
V.N. Karazin Kharkiv National University	Ukraine
ВСП «Рівненський технічний фаховий коледж Національного університету водного господарства та природокористування»	Україна
Вінницький національний технічний університет	Україна
ВСП «Одеський технічний фаховий коледж ОНТУ»	Україна
ВТЕІ КНТЕУ	Україна
ДВНЗ "Український державний хіміко-технологічний університет"	Україна
Державна наукова установа «Науково-практичний центр профілактичної та клінічної медицини» Державного управління справами	Україна
Дніпровський національний університет імені Олеся Гончара	Україна
Донбаська державна машинобудівна академія	Україна
Донецький національний технічний університет	Україна
Економіко-технологічний інститут ім. Роберта Ельворті	Україна
Івано-Франківський національний технічний університет нафти і газу	Україна
Інститут кібернетики імені В.М. Глушкова НАН України	Україна
Інститут проблем штучного інтелекту НАН України та МОН України	Україна
Інститут транспортних систем та технологій Національної академії наук України	Україна
Комунальна установа Сумська спеціалізована школа I-III ступенів №25	Україна
Криворізький національний університет	Україна
Львівський торговельно-економічний університет	Україна
Міжнародний європейський університет	Україна
Міжнародний науково-навчальний центр інформаційних технологій та систем НАН	Україна
Національний аерокосмічний університет ім. М. Є. Жуковського "ХАІ"	Україна
Національний технічний університет «Харківський політехнічний інститут»	Україна
Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"	Україна

Національний університет «Львівська політехніка»	Україна
Національний університет «Одеська морська академія»	Україна
Національний університет «Одеська політехніка»	Україна
Національний університет біоресурсів і природокористування України	Україна
Одеський національний технологічний університет	Україна
Одеський національний університет імені І.І. Мечникова	Україна
Сумський державний педагогічний університет імені А.С.Макаренка	Україна
Український державний університет науки і технологій	Україна
Український державний хіміко-технологічний університет	Україна
Університет митної справи та фінансів	Україна
Харківський національний університет радіоелектроніки	Україна
Херсонська державна морська академія	Україна
Чорноморський національний університет імені Петра Могили	Україна

УДК 004.891

МЕТОДИКА ЧАСТОТНОГО АНАЛІЗУ ТЕКСТУ ЗА ДОПОМОГОЮ АЛГОРИТМА COUNT-MIN SKETCH

Акішев О.О., Арсірій О.О. (akishev.oleksii@gmail.com , e.arsiriy@gmail.com)

Національний університет «Одеська Політехніка» (Україна)

Розглянуто недоліки хеш-таблиць щодо зберігання великих потокових текстових даних. Для їх подолання запропоновано використання модифікованого хешування даних на основі імовірнісного алгоритму Count-Min-Sketch. Розроблено методику його використання для зберігання великих текстових даних та їх подальшого частотного аналізу з метою отримання списку ключових слів. Для реалізації вибраного алгоритму використовувалась мова програмування Java, а в якості середовища розробки - IntelliJ IDEA 2022.

Вступ. Частотний аналіз тексту є одним із методів обробки тексту на природній мові (Natural Language Processing, NLP), що спрямовані на вилучення структурованої інформації із неструктурованих текстів. При цьому корисним результатом може бути пошук ключових слів та словосполучень, так званих основних понять, які найчастіше зустрічаються в аналізованому тексті. Отримання частотного списку основних слів тексту може бути корисним в подальшому для вирішення задач кластеризації та класифікації великих об'ємів текстової інформації

Мета роботи. Метою роботи є зниження ресурсоемності отримання списку ключових слів (характерних ознак) із великих тестових даних за рахунок розробки методики частотного аналізу текстів на основі алгоритму Count-Min-Sketch [1,2,3].

Основна частина роботи. При розробці методики важливим є вирішення проблеми представлення та збереження даних з вхідного потоку. Зазвичай для отримання частоти будь-яких потокових даних використовуються механізм хеш-таблиць, який дозволяє зберігати хеш-значення та легко їх отримувати за часовою складністю $O(1)$. Але в випадку отримання великих потокових вхідних використання хеш-таблиць для їх зберігання з метою подальшого частотного аналізу призводить до суттєвого підвищення часової складності O , а гіршому випадку до її лінійної залежності $O(n)$, де n – розмір тексту. Для запобігання неефективного використання оперативної пам'яті для зберігання великих даних та часу для їх отримання з метою подальшого частотного аналізу запропоновано використання алгоритму Count-Min-Sketch. Його реалізація дозволяє зберігати не повні значення потоку даних у вигляді хеш-таблиці, а побудувати двовимірний масив, так званий Sketch (рис. 1), який є масивом лічильників, які відображають кількість елементів в вхідному потоці і набором хеш-функцій. Точніше, Sketch створюється як послідовність рядків, де кожний елемент із вхідного потоку хешується першою хеш-функцією в перший рядок, другою хеш-функцією у другий рядок і так далі. При цьому нарощується значення лічильника у відповідній позиції відповідного рядка.

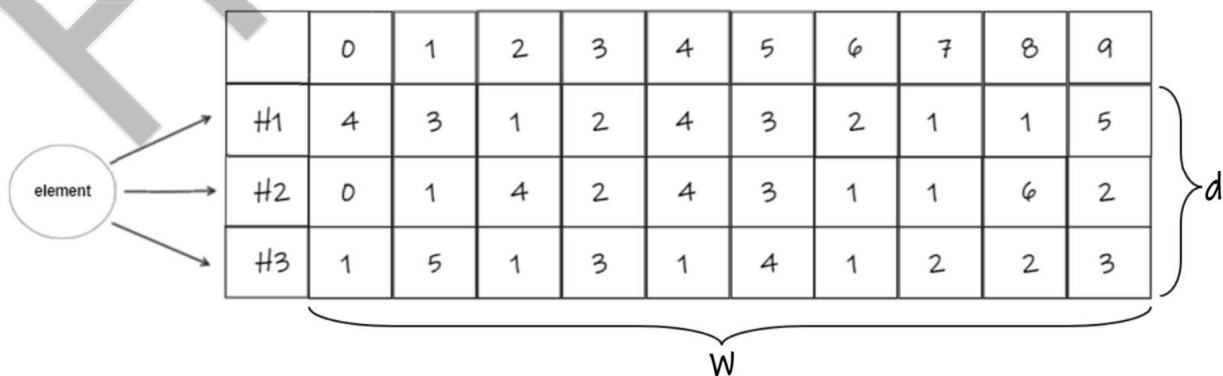


Рис. 1 –Вигляд двовимірного вихідного масиву (Sketch) за алгоритмом Count-Min-Sketch

Для отримання частоти елемента вхідного потоку необхідно виконати його ітеративне хешування та зіставлення зі значеннями відповідних лічильників в кожному рядку. Згідно з процедурою додавання вхідного елемента за алгоритмом Count-Min-Sketch він представляє собою певне число в кожному рядку Sketch-таблиці, тому в якості частоти обираємо мінімальне значення з набору чисел який ми отримуємо в результаті зіставлення всіх хеш-функцій. Це буде комірка, яка мала найменшу кількість конфліктів хешу з іншими комірками. Наприклад, якщо для елемента «А» по кожній з хеш-функцій ми отримуємо множину значень $A=\{3, 2, 5\}$, тоді його частотою вважається найменше число яке знаходиться в множині елемента «А», тобто «2».

Якщо враховувати той факт, що алгоритм Count-Min-Sketch є імовірнісним необхідно визначитись з необхідною точністю частотного аналізу. При цьому необхідно враховувати, що чим більше хеш-функцій використано для побудови Sketch-таблиці та взагалі чим більша її розмірність, тим менше виникає колізій, тобто менший відсоток помилки. Проте з точки зору ресурсоемності неефективно створювати та використовувати велику за розміром Sketch-таблицю, яка буде заповнена лише на половину. Отже, для проведення ефективного частотного аналізу за алгоритмом Count-Min-Sketch в залежності від потенційної кількості елементів вхідного потоку – «n», необхідно визначити приблизну ширину «w» та глибину «d» Sketch-таблиці. При цьому потрібно виконати наступні кроки:

1. Визначити, скільки приблизно кількість елементів вхідного потоку - n.

2. Визначити відсоток помилки «error» від суми усіх лічильників для підрахунку кількості елементів.

3. Визначити імовірність, тобто оцінити точність «δ» роботи алгоритму.

4. Визначити кількість колонок/лічильників які представлені шириною «w».

5. Визначити кількість рядків/хеш-функцій які представлені глибиною «d».

Визначимо наступну постановку задачі частотного аналізу тексту. Вхідний набір даних зчитується з текстового файлу, в даному файлі зберігається текст художньої книги. Завдання полягає у тому щоб знайти певну кількість найчастіше повторюваних слів у тексті та відсортувати їх у порядку спадання, від елемента який зустрічався більше всіх до можна представити послідовністю наступних кроків:

1. Зчитуючи дані з текстового файлу створюємо список або масив, за допомогою знаків пунктуації.

2. Видаляємо деякі слова, які є очевидними як артиклі, займенники, та інші.

3. Вводимо заздалегідь визначені змінні, а саме «δ», «w», «d», а також визначаємо бажану кількість вихідних елементів, тобто ключових слів.

4. Створення порожньої робочої Sketch-таблиці.

5. Побудова результуючої Sketch-таблиці за допомогою ітеративного хешування відповідного елемента вхідного масиву для отримання позиції лічильника та збільшення його значення на одиницю.

6. Використання результуючої Sketch-таблиці для отримання частоти елементів вхідного масиву у вигляді відповідної множини значень лічильників. Визначення позиції лічильників у рядках Sketch-таблиці за допомогою ітеративного хешування.

Для порівняння точності частотного аналізу за алгоритмом Count-Min-Sketch виконувалась побудова класичної хеш-таблиці для збереження елементів вхідного текстового файлу. На рисунку 2 показана результуюча таблиця з порівняльними характеристиками частот ключових слів тексту отриманими за двома підходами. В таблиці наведено точне значення кількості слів (стовпчик «freq_ref») з відповідною назвою (стовпчик «word») в аналізованому тексті, яке отримано за допомогою побудови класичної хеш-таблиці та імовірнісне значення кількості слів (стовпчик «freq_approx»), яке отримано за алгоритмом Count-Min-Sketch. Останній стовпчик «error %» показує різницю між значеннями «freq_ref» та «freq_approx» у відсотковому співвідношенні. Використовуючи вбудовані методи бібліотек Java можна порівняти кількість витрачених ресурсів, а саме пам'яті на збереження та сортування елементів, за допомогою алгоритму Count-Min-Sketch та класичних хеш-

таблиць. Отже, на збереження та сортування 16 тисяч елементів досліджуваним алгоритмом Count-Min-Sketch було використано - 0.08 МБ пам'яті, в той же самий час звичайній хеш таблиці було потрібно - 2.67 МБ.

word	freq_ref	freq_approx	error %
Buck	360	366	1.67
dogs	118	144	22.03
Thornton	102	113	10.78
Gutenberg	84	102	21.43
Project	84	91	8.33
Spitz	65	89	36.92
dog	76	88	15.79
sled	63	85	34.92
men	69	83	20.29
Fran	60	75	25.0

Рис. 2 – Результати порівняння використання алгоритму Count-Min Sketch та збереження елементів у класичній хеш-таблиці для частотного аналізу текстів

Висновки. Таким чином розроблено методику частотного аналізу текстів з використанням алгоритму Count-Min Sketch. Методику реалізовано за допомогою мови програмування Java в середовище IntelliJ IDEA 2022 та апробовано на прикладі частотного аналізу тексту «*Jack London, The Call of the Wild*». Результати наведено в таблиці (рис. 2). В зв'язку з наведеними в тексті особливостями алгоритму Count-Min-Sketch можна гарантувати, що імовірнісна оцінка «freq_approx» або дорівнює значенню точної оцінки «freq_ref», або перевищує його за рахунок наявності так званих хибно позитивних результатів. Але і іншого боку можна завжди гарантувати що ця імовірнісна оцінка ніколи не буде меншою ніж точна частота слова в тексті. Що стосується ресурсоемності обох підходів, то використання імовірнісного підходу дає перевагу по об'єму оперативної пам'яті в 33 рази. Це факт свідчить про перспективність використання імовірнісних алгоритмів та структур даних для частотного аналізу великих обсягів текстової інформації.

Список використаної літератури

- [1] “Probabilistic Data Structures and Algorithms for BigData applications,” Andrii Gakhov. [Online]. Available: https://books.google.fr/books?id=PWOHDwAAQBAJ&printsec=frontcover&redir_esc=y#v=onepage&q&f=false [Accessed: October 08, 2022]
- [2] “The Count-Min Sketch and its Applications,” Graham Cormode, S. Muthukrishnan. [Online]. Available: <http://dimacs.rutgers.edu/~graham/pubs/papers/cm-full.pdf> [Accessed: October 08, 2022]
- [3] “Big Data with Sketchy Structures,” Karan Shukla. [Online]. Available: <https://towardsdatascience.com/big-data-with-sketchy-structures-part-1-the-count-min-sketch-b73fb3a33e2a> [Accessed: October 08, 2022]

XV МІЖНАРОДНА НАУКОВО-ПРАКТИЧНА КОНФЕРЕНЦІЯ

**«ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ І
АВТОМАТИЗАЦІЯ – 2022»**

**20 - 21 ЖОВТНЯ 2022 р.
м.Одеса**

XV INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE

**«INFORMATION TECHNOLOGIES AND
AUTOMATION– 2022»**

**OCTOBER 20 - 21, 2022
Odessa**

Збірник включає доповіді учасників конференції. Тези доповідей публікуються у вигляді, в якому вони були подані авторами.

Відповідальність за зміст і форму подачі матеріалу несуть автори статей.

The collection includes reports of conference participants. Abstracts are published in the form in which they were submitted by the authors.

The authors of the articles are responsible for the content and form of submission of the material.

Редакційна колегія: Котлик С.В., Корнієнко Ю.К., Ломовцев П.Б.

Комп'ютерний набір і верстка: Соколова О.П.

Відповідальний за випуск: Котлик С.В.