

НАУКОВІ ДОСЛІДЖЕННЯ МЕТРИКИ ХЕММІНГА: ВІД ТЕОРІЇ КОДУВАННЯ ДО ШТУЧНОГО ІНТЕЛЕКТУ

Відстань Хеммінга – це метрика, яка визначає кількість позицій, у яких два рядки однакової довжини відрізняються один від одного. Вперше її запропонував американський математик Річард Хеммінг у 1940-х роках у контексті задач кодування та виправлення помилок у передаванні даних. Первинно відстань Хеммінга використовувалася у теорії кодування для розробки алгоритмів виявлення та виправлення помилок під час передачі інформації каналами зв'язку. З часом сфери застосування відстані Хеммінга значно розширилися. Сьогодні її використовують у криптографії та безпеці (аналіз стійкості алгоритмів шифрування, виявлення вразливостей і побудова апаратних систем захисту), алгоритмах та оптимізації (як критерій подібності у задачах класифікації, хешування та евристичного пошуку), біометрії та розпізнавання образів (зіставлення шаблонів у системах ідентифікації), біоінформатиці та генетиці (порівняння нуклеотидних послідовностей, моделювання генетичних варіацій), а також екологічних науках (моніторингу довкілля за принципом «подібність/відмінність» у часових рядах або просторових структурах) [1-7]. У природничих науках відстань Хеммінга використовується як частина гібридних моделей для прогнозування та аналізу якості компонентів довкілля [8-10]. Таким чином, відстань Хеммінга є універсальним математичним інструментом, що поєднує фундаментальні принципи інформатики з широким спектром прикладних досліджень у природничих, технічних та медико-біологічних науках. Для візуалізації основних напрямів досліджень із використанням відстані Хеммінга, було проведено бібліометричний аналіз за даними наукометричної бази Scopus. У результаті пошуку за ключовим словом «hamming distance» виявлено 9962 наукових публікацій, що охоплюють період з 60-тих років XX століття й до сьогодення (рис. 1).

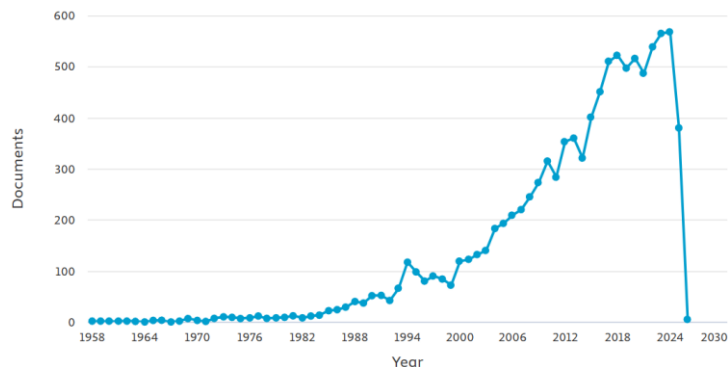


Рис. 1. Результати пошуку за ключовим словом «hamming distance» у наукометричній базі Scopus

За допомогою програмного забезпечення VOSviewer побудовано бібліографічну хмару ключових термінів, яка відображає міждисциплінарність застосування даної метрики (рис. 2).

Виділено 7 кластерів застосування відстані Хеммінга: 1) червоний – теорія кодування, що застосовується у передачі даних, телекомунікації, виявленні та виправленні помилок, цифрових системах зберігання (ключові слова кластеру: «codes (symbols)» (коди (символи)), «block codes» (блокові коди), «linear codes» (лінійні коди), «decoding» (декодування), «bit error rate» (ймовірність бітової помилки)); 2) синій – криптографія та апаратна безпека, що стосується криптосистем, безпеки апаратних пристроїв, захисту від атак (ключові слова: «cryptography» (криптографія), «hardware security» (апаратна безпека), «side-channel attack» (атаки через побічні канали), physical unclonable functions (PUF) (фізично неклоновані функції)); 3) жовто-зелений – алгоритми та оптимізація, що застосовуються для порівняння структур даних, хешування, машинного навчання, пошукових й класифікаційних алгоритмів (ключові слова: «hash functions» (хеш-функції), «semantics» (семантика), «learning algorithms» (алгоритми навчання), «optimization» (оптимізація), «problem solving» (розв'язання задач)); 4) фіолетовий – біометрія та обробка зображень, що застосовується у системах ідентифікації (відбитки пальців, райдужка), розпізнавання образів, відеоаналітиці (ключові слова: «biometrics» (біометрія), «image processing» (обробка зображень), «pattern recognition» (розпізнавання образів), «iris recognition» (розпізнавання райдужки)); 5) зелений – біоінформатика та генетика, що застосовується для порівняння геномних послідовностей, аналізу SNP, обчислювальній біології, медичних дослідженнях (ключові слова: «bioinformatics» (біоінформатика), «computational biology» (обчислювальна біологія), «genetics» (генетика), «DNA» (ДНК), «polymorphism» (поліморфізм)); 6) помаранчевий – ухвалення рішень і нечіткі множини, що застосовується у теорії рішень, нечіткій логіці, експертних системах, багатокритеріальному аналізі («decision making» (ухвалення рішень), «fuzzy sets» (нечіткі множини), «aggregation operator» (оператори агрегування), «uncertainty» (невизначеність)); 7) бірюзовий – еволюційні та евристичні алгоритми, що використовуються у штучному інтелекті, еволюційних обчисленнях, оптимізації складних задач (ключові слова: «genetic algorithms» (генетичні алгоритми), «approximation algorithms» (алгоритми наближення), «heuristics» (евристики), «population statistics» (статистика популяцій)).



Рис. 2. Бібліометрична карта ключових напрямів досліджень відстані Хеммінга за публікаціями Scopus (n = 9962)

Отже, відстань Хеммінга є універсальним інструментом, що поєднує класичні математичні властивості з широкими міждисциплінарними застосуваннями – від кібербезпеки до екологічного моніторингу. Сучасні дослідження демонструють тенденцію до інтеграції цієї метрики з алгоритмами штучного інтелекту, криптографією та моделями прийняття рішень у складних системах.

Використана література:

1. Cachet, C., Ahmad, S., Demarest, L., Hamlin, A., & Fuller, B. W. (2022). Proximity searchable encryption for the iris biometric. *Proceedings of the ACM Conference*, 1004–1018. <https://doi.org/10.1145/3488932.3497754>.
2. Hofheinz, D., Jager, T., & Knapp, E. (2012). Waters signatures with optimal security reduction. In *Lecture Notes in Computer Science* (Vol. 7293, pp. 66–83). Springer. https://doi.org/10.1007/978-3-642-30057-8_5.
3. Lin, Z. (2025). A parametric design method for landscape garden layout based on clustering algorithm. *International Journal of High Speed Electronics and Systems*, 34(4), 2540264. <https://doi.org/10.1142/S0129156425402645>.
4. Meng, P., Jia, S., & Li, Q. (2021). Recognition of video time watermark based on water-wave connected domain segmentation algorithm. *ICCT 2021 Proceedings*, 1123–1127. <https://doi.org/10.1109/ICCT52962.2021.9658031>.
5. Pandey, R. K., & Das, T. K. (2023). Anomaly detection for industrial control networks using Hamming distance. In *Lecture Notes in Networks and Systems* (Vol. 521, pp. 280–290). Springer. https://doi.org/10.1007/978-3-031-13150-9_24.
2. Pandey, R. K., & Das, T. K. (2025). Anomaly detection in cyber-physical systems using actuator state transition model. *International Journal of Information Technology* (Singapore), 17(3), 1509–1521. <https://doi.org/10.1007/s41870-024-02128-x>.
3. Qiu, L., Chen, W., Li, F., Liu, S., Wang, X., & Li, L. (2022). Fast Hash_LBP moving target detection algorithm based on Hamming distance constraint in complex background. *Acta Photonica Sinica*, 51(9), 0910003. <https://doi.org/10.3788/gzxb20225109.0910003>.
4. Si, X., & Zhang, Y. (2016). Quality assessment of surface water environment based on improved grey correlation method. *Conference Proceedings*, 175–180.
5. Valerko, R. A., Herasymchuk, L. O., & Romanchuk, L. D. (2024). Modeling drinking water quality using the “Hamming distance” method. In *Ecological state of the environment and rational nature management in the context of sustainable development: Proceedings of the 7th International Scientific and Practical Conference* (Kherson, Ukraine, October 24–25, 2024) (pp. 74–77). Odessa: Oldi+.
6. Vijayalakshmi, C., & Pushpa, M. (2025). Optimizing non-linear autoregressive networks with Bird Sea Lion algorithms for effective rainfall forecasting. *Earth Science Informatics*, 18(3), 279. <https://doi.org/10.1007/s12145-025-01768-2>.