УДК 004.421.2

В.М. Липкань
Н.В. Добровольська

ХЕШУВАННЯ ЯК МЕТОД ОРГАНІЗАЦІЇ ТА ПОШУКУ ДАНИХ

Вінницький національний технічний університет

Анотація

У доповіді розглянуто хешування як фундаментальний метод швидкого доступу до даних. Хеш-функція трактується як відображення множини ключів на множину індексів таблиці. Проаналізовано проблему колізій та методи їх розв'язання – ланцюжки і відкриту адресацію. Показано, що за рівномірної хеш-функції основні операції виконуються в середньому за сталий час. Окремо розглянуто криптографічне хешування та сфери практичного застосування.

Ключові слова: хеш-функція, хеш-таблиця, колізія, відкрита адресація, складність, криптографічне хешування.

Abstract

The report considers hashing as a fundamental method of fast data access. A hash function is treated as a mapping of a set of keys onto a set of table indices. The collision problem and methods of its resolution – chaining and open addressing – are analysed. It is shown that with a uniform hash function the main operations are performed on average in constant time. Cryptographic hashing and areas of practical application are also discussed.

Keywords: hash function, hash table, collision, open addressing, complexity, cryptographic hashing.

Вступ

Однією з найпоширеніших задач в обчислювальній техніці є швидкий пошук, додавання та видалення даних. Лінійний пошук у невпорядкованому масиві потребує в середньому перегляду половини елементів, а бінарний пошук вимагає підтримки впорядкованості. Хешування дозволяє виконувати ці операції в середньому за сталий час O(1), що робить його основою багатьох ефективних структур даних. Ідея методу полягає в перетворенні ключа на індекс у масиві фіксованого розміру за допомогою спеціальної функції.

Хеш-функція та вимоги до неї

Хеш-функція h(k) відображає множину ключів K на множину індексів $\{0, 1, \ldots, m-1\}$, де m – розмір хеш-таблиці. Формально $h: K \to \{0, \ldots, m-1\}$. Оскільки множина можливих ключів зазвичай значно більша за розмір таблиці, різні ключі можуть відображатися в одну й ту саму комірку – виникає колізія. До «хорошої» хеш-функції висувають вимоги рівномірності розподілу, детермінованості, швидкості обчислення та залежності результату від усіх частин ключа.

Найпростішим є метод ділення: $h(k) = k \bmod m$, де розмір m доцільно обирати простим числом. У методі множення обчислюють $h(k) = \lfloor m \cdot (k \cdot A \bmod 1) \rfloor$, де A – стала в інтервалі (0, 1). Для рядків застосовують поліноміальну схему, що враховує коди всіх символів та їхній порядок.

Колізії та методи їх розв'язання

Колізія виникає, коли $h(k1) = h(k2)$ для різних ключів $k1 \neq k2$. Повністю уникнути колізій неможливо, тому ключовим є їх ефективна обробка. У методі ланцюжків кожна комірка таблиці містить список усіх елементів з однаковим значенням хешу. У відкритій адресації всі елементи зберігаються безпосередньо в таблиці, а при зайнятості комірки за певним правилом пробуються інші:

$$h(k, i) = (h(k) + i) \bmod m$$ – лінійне пробування;

$$h(k, i) = (h(k) + c1 \cdot i + c2 \cdot i^2) \bmod m$$ – квадратичне пробування;

$$h(k, i) = (h1(k) + i \cdot h2(k)) \bmod m$$ – подвійне хешування.

Важливою характеристикою є коефіцієнт заповнення $\alpha = n/m$. Зі зростанням α збільшується ймовірність колізій, тому при перевищенні порогового значення (зазвичай $\alpha \approx 0{,}7$) виконують рехешування – створення більшої таблиці з перерозподілом елементів.

Аналіз часової складності

Ефективність операцій у хеш-таблиці залежить від якості хеш-функції та коефіцієнта заповнення. У таблиці 1 наведено асимптотичну складність основних операцій.

Таблиця 1 – Складність операцій у хеш-таблиці

Операція	**Середній випадок**	**Найгірший випадок**
Пошук	O(1)	O(n)
Вставлення	O(1)	O(n)
Видалення	O(1)	O(n)

Найгірший випадок O(n) виникає, коли всі ключі потрапляють в одну комірку. За умови рівномірної хеш-функції та контрольованого коефіцієнта заповнення такий сценарій є малоймовірним, тому на практиці хеш-таблиці забезпечують майже сталий час роботи.

Криптографічне хешування

Окремий клас становлять криптографічні хеш-функції (SHA-256 та інші). До них висувають додаткові вимоги: стійкість до прообразу (за хешем неможливо відновити вихідні дані), стійкість до колізій та лавинний ефект, за якого мінімальна зміна входу кардинально змінює результат. Такі функції застосовують для перевірки цілісності даних, зберігання паролів із «сіллю», електронного цифрового підпису та в технології блокчейн.

Висновок

Хешування є потужним і універсальним методом, що забезпечує швидкий доступ до даних у середньому за сталий час. Якість роботи хеш-таблиці визначається передусім вдалим вибором хеш-функції та стратегією розв’язання колізій. Метод ланцюжків і відкрита адресація мають свої переваги залежно від характеру задачі. Криптографічне хешування розширює застосування методу на задачі інформаційної безпеки, завдяки чому хешування залишається одним із базових інструментів сучасних алгоритмів та структур даних.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Кормен Т., Лейзерсон Ч., Рівест Р., Стайн К. Алгоритми: побудова та аналіз. 3-тє вид. Київ: К.І.С., 2019. 1288 с.
2. Knuth D. E. The Art of Computer Programming. Volume 3: Sorting and Searching. 2nd ed. Reading, Massachusetts: Addison-Wesley, 1998. 780 p.
3. Sedgewick R., Wayne K. Algorithms. 4th ed. Upper Saddle River, NJ: Addison-Wesley, 2011. 952 p.
4. Aho A. V., Hopcroft J. E., Ullman J. D. Data Structures and Algorithms. Reading, Massachusetts: Addison-Wesley, 1983. 427 p.

Липкань Віктор Михайлович – студент групи 1КІ-25МС, факультет інформаційних технологій та комп’ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, e-mail: lipkan02@gmail.com.

Науковий керівник: ***Добровольська Наталія Вікторівна*** – кандидат педагогічних наук, доцент, кафедра обчислювальної техніки, Вінницький національний технічний університет, м. Вінниця.

Viktor Mykhailovych Lypkan – student of group 1KI-25MS, Faculty of Information Technology and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: lipkan02@gmail.com.

Supervisor: ***Dobrovolska Natalia V.*** – Candidate of Pedagogical Sciences, Associate Professor, Department of Computer Engineering, Vinnytsia National Technical University, Vinnytsia.