



Prefix Array



Longest Common Prefix (LCP)

จัดทำโดย กลุ่ม อีซูอีซู



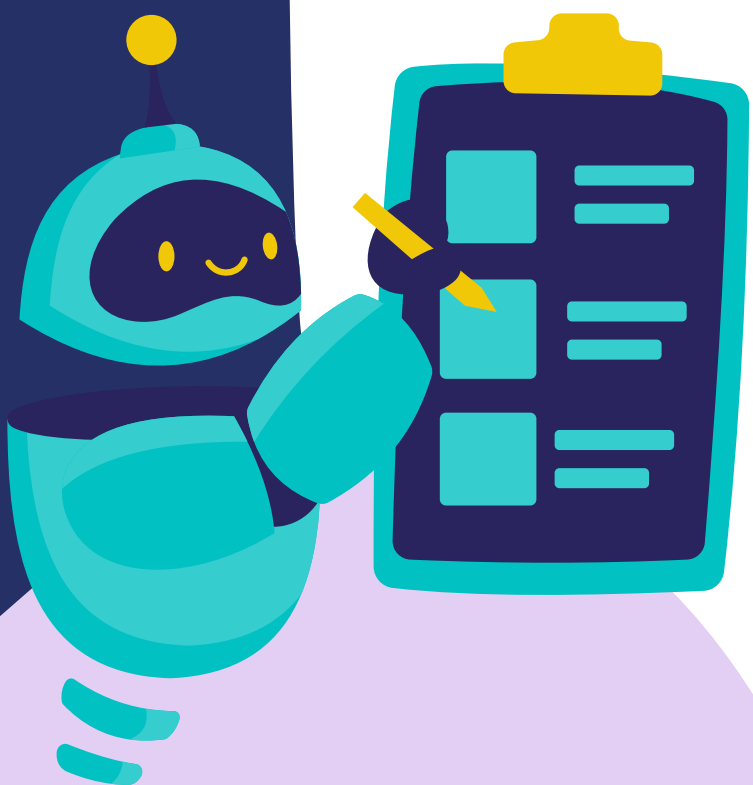


Prefix Array

Prefix Array เป็นแนวคิดที่ใช้ในหลายอัลกอริทึม เช่น การประมวลผลข้อความ (String Processing) หรือการคำนวณในโครงสร้างข้อมูล โดย **Prefix Array** มีจุดประสงค์เพื่อจัดเก็บข้อมูลบางส่วนหรือผลลัพธ์ของข้อมูลต้นฉบับในลักษณะที่ช่วยให้เราสามารถเข้าถึงหรือคำนวณข้อมูลได้เร็วขึ้น

หลักการทำงาน

หลักการทำงานของ **Longest Common Prefix (LCP)** คือ การหาส่วนเริ่มต้น (prefix) ที่เหมือนกันของข้อความ (string) ใน array ทั้งหมด โดยทำการเปรียบเทียบตั้งแต่ตัวอักษรแรกของแต่ละคำ ไปเรื่อย ๆ จนกว่าจะเจอความแตกต่าง หรือไม่มี prefix ร่วมกันเลย



LCP Array

การเปรียบเทียบ คำว่า String AAAAAAAAA\$

1. \$ กับ A\$ → ไม่มี prefix ที่เหมือนกัน → LCP = 0
2. A\$ กับ AA\$ → prefix เหมือนกันคือ "A" → LCP = 1
3. AA\$ กับ AAA\$ → prefix เหมือนกันคือ "AA" → LCP = 2
4. AAA\$ กับ AAAA\$ → prefix เหมือนกันคือ "AAA" → LCP = 3
5. AAAAA\$ กับ AAAAAA\$ → prefix เหมือนกันคือ "AAAAA" → LCP = 4
6. AAAAAA\$ กับ AAAAAAA\$ → prefix เหมือนกันคือ "AAAAAA" → LCP = 5
7. AAAAAAA\$ กับ AAAAAAAA\$ → prefix เหมือนกันคือ "AAAAAAA" → LCP = 6

ยกตัวอย่างการทำงาน

1) Input : strs = ["flower", "flow", "flight"]

2) กำหนด prefix เริ่มต้นเป็นคำแรก :

prefix = "flower"

3) เปรียบเทียบกับ string ที่เหลือ :

- เปรียบเทียบกับ "flow" :

- "flower" ไม่เริ่มต้นด้วย "flow" → ลด prefix ลงเป็น "flow"

- เปรียบเทียบกับ "flight" :

- "flow" ไม่เริ่มต้นด้วย "flight" → ลด prefix ลงเป็น "flo", "fl"

- "fl" เป็น prefix ที่ตรงกัน

4) จบการทำงาน :

Longest Common Prefix = "fl"



ประโยชน์ของ LCP

การหา LCP มีประโยชน์ในหลายกรณี เช่น:

การจัดกลุ่มข้อความ (Clustering Strings) :

- ใช้หา similarity ระหว่างข้อความ เช่น การจัดหมวดหมู่ข้อมูล

การค้นหาในฐานข้อมูล :

- ใช้สำหรับแนะนำคำค้นหาที่มี prefix เหมือนกัน

ในอัลกอริทึม String Matching :

- เช่น ใช้ในการคำนวณในอัลกอริทึม KMP (Knuth-Morris-Pratt) เพื่อค้นหาข้อความย่อยในข้อความใหญ่

ประสิทธิภาพของอัลกอริทึม

- เวลาในการทำงาน (Time Complexity):

$O(n \times m)$ Best case - $O(\minLen \times n)$, Worst case - $O(\minLen \times n)$

โดย n คือจำนวนคำในอาร์เรย์ และ m คือความยาวเฉลี่ย
ของคำแต่ละคำเพราะเราต้องเปรียบเทียบคำทั้งหมดกับ prefix

- พื้นที่ในการทำงาน (Space Complexity):

$O(1)$

ไม่ต้องใช้พื้นที่เพิ่มเติมยกเว้นตัวแปร prefix



สรุป

Longest Common Prefix

- เป็นการหาส่วนเริ่มต้นที่เหมือนกันของข้อความทั้งหมด
- มีประโยชน์ในหลายกรณี เช่น การจัดกลุ่มข้อความ หรือ ค้นหาคำที่มีความคล้ายกัน
- อัลกอริทึมทำงานโดยการลดขนาด prefix จนเจอ คำที่เหมือนกันทั้งหมด หรือไม่มี prefix ร่วมกันเลย

จัดทำโดย

นาย พิพรรณ คลายสุข 6710301003

นาย ธนัท เนตรยอง 6720301003

นาย สิริภพ เสวี่ยมศักดิ์ 6710301036

นาย ชลนภัส รักษ์เจริญ 6710301051

จบการนำเสนอ
Thank you

