



Longest Repeated Substring (LRS)

LONGEST REPEATED SUBSTRING (LRS) คือปัญหาที่เกี่ยวข้องกับการค้นหาชุดของอักขระที่ยาวที่สุดซึ่งปรากฏซ้ำกันมากกว่าหนึ่งครั้งในสตริงที่กำหนด ตัวอย่างเช่น ในสตริง BANANA คำตอบคือ ANA ซึ่งเป็นชุดของอักขระที่ซ้ำกันและยาวที่สุด



การประยุกต์ใช้งาน

LONGEST REPEATED SUBSTRING มีประโยชน์ในหลายด้าน เช่น:

ชีวสารสนเทศศาสตร์ (BIOINFORMATICS): ใช้ในการ
วิเคราะห์ DNA เพื่อค้นหาลำดับที่ซ้ำกัน

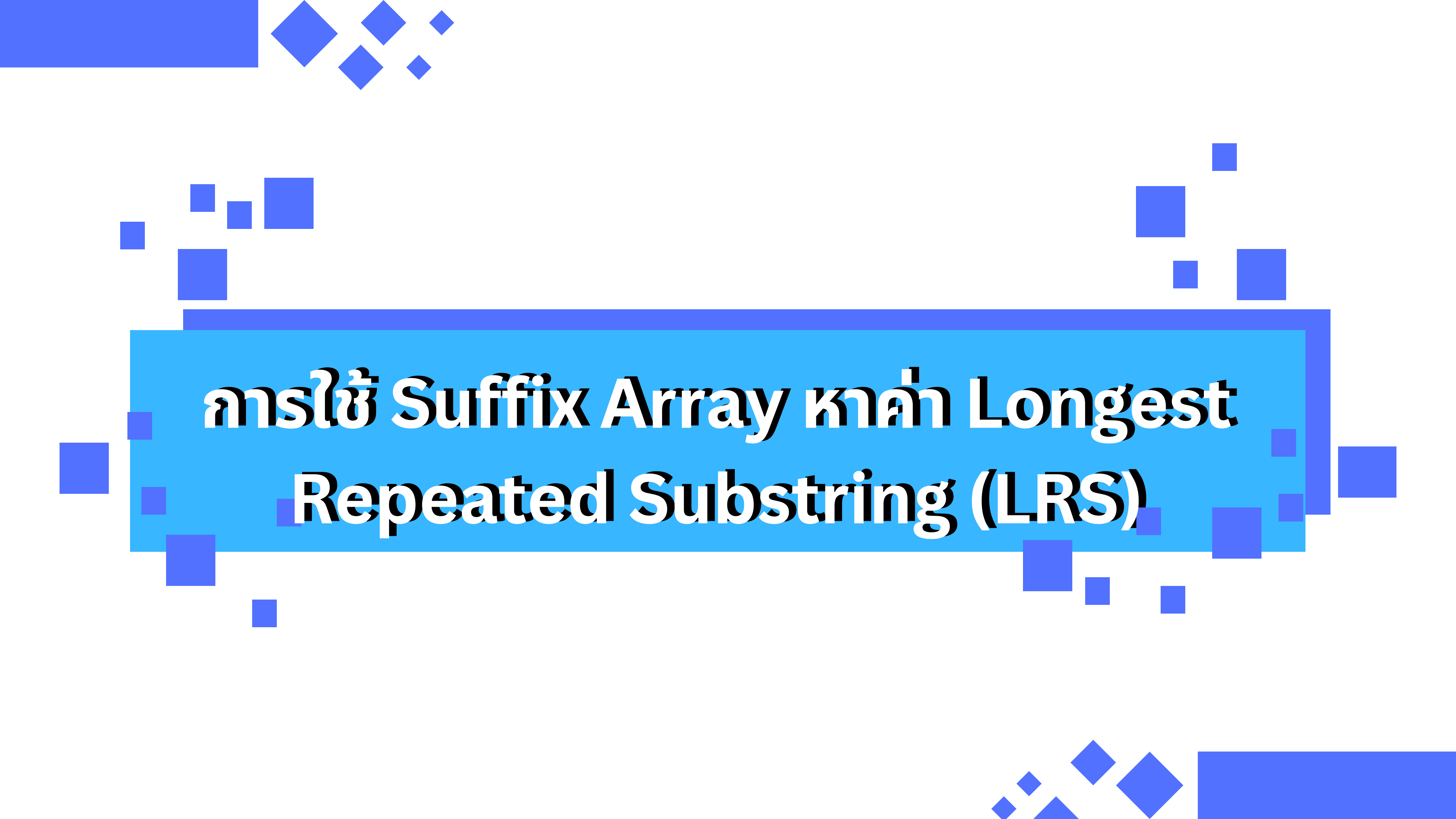
การบีบอัดข้อมูล (DATA COMPRESSION): ใช้ค้นหารูป
แบบที่ซ้ำกันเพื่อลดขนาดข้อมูล

วิเคราะห์ข้อความ (TEXT ANALYSIS): ใช้ค้นหารูปแบบ
ซ้ำในเอกสารหรือข้อความ

การป้องกันการลอกเลียนแบบ: ใช้ตรวจสอบข้อความหรือ
โค้ดที่ซ้ำกันในหลายแหล่งที่มา

ข้อมูลควรรู้

- SUFFIX ARRAY คือ อาร์เรย์ที่เก็บตำแหน่งเริ่มต้นของ SUFFIX ของสตริงที่เรียงตามลำดับตัวอักษร LEXICOGRAPHICAL ORDER (พจนานุกรม)
- LCP ARRAY คือ อาร์เรย์ที่เก็บความยาวของ PREFIX ที่เหมือนกันระหว่าง SUFFIXES ที่เรียงติดกันใน SUFFIX ARRAY
- LONGEST REPEATED SUBSTRING (LRS) คือ ค่าที่มากที่สุดภายใน LCP ARRAY
- SUFFIX คือ ส่วนที่เหลือของสตริงตั้งแต่ตำแหน่งใด ๆ ไปจนถึงตัวสุดท้าย (ตัดตัวแรก)
- PREFIX คือ ส่วนของสตริงตั้งแต่ตัวแรกไปจนถึงตำแหน่งใด ๆ ในสตริง (ตัดตัวสุดท้าย)

The background features a light blue gradient with various blue geometric shapes, including squares and diamonds of different sizes, scattered across the frame. A central blue rectangular box contains the text.

การใช้ Suffix Array ค้นหา Longest Repeated Substring (LRS)

ขั้นตอนการสร้าง SUFFIX ARRAY:

1. สร้าง SUFFIX ของสตริงพร้อมตำแหน่งเริ่มต้น:

- "BANANA" (INDEX 0)
- "ANANA" (INDEX 1)
- "NANA" (INDEX 2)
- "ANA" (INDEX 3)
- "NA" (INDEX 4)
- "A" (INDEX 5)

2. เรียง SUFFIX ตามพจนานุกรม:

- "A" (INDEX 5)
- "ANA" (INDEX 3)
- "ANANA" (INDEX 1)
- "BANANA" (INDEX 0)
- "NA" (INDEX 4)
- "NANA" (INDEX 2)

3. เก็บเฉพาะตำแหน่งเริ่มต้นของ SUFFIX:

- SUFFIX ARRAY = [5, 3, 1, 0, 4, 2]

วิธีสร้าง LCP ARRAY:

2

1. เปรียบเทียบแต่ละคู่ SUFFIX ที่เรียงกันใน SUFFIX ARRAY:

- "A" กับ "ANA" → LCP = 1 ("A")
- "ANA" กับ "ANANA" → LCP = 3 ("ANA")
- "ANANA" กับ "BANANA" → LCP = 0 (ไม่มีตัวอักษรเหมือนกัน)
- "BANANA" กับ "NA" → LCP = 0 (ไม่มีตัวอักษรเหมือนกัน)
- "NA" กับ "NANA" → LCP = 2 ("NA")

2. ได้ LCP ARRAY = [1, 3, 0, 0, 2, 0]

หาค่า LONGEST REPEATED SUBSTRING (LRS)

- LRS คือค่าที่มากที่สุดใน LCP ARRAY
- ในตัวอย่างนี้ $\text{MAX}(\text{LCP}) = 3$ ซึ่งหมายถึงซับสตริงที่ซ้ำกันยาวที่สุดมีความยาว 3
- จาก SUFFIX ARRAY ตำแหน่งของซับสตริงนี้เริ่มที่ $\text{INDEX SUFFIX}[1] = 3$ (ใน "ANANA")
- ซับสตริงที่ซ้ำกันคือ $s[3:3+3] = \text{"ANA"}$

สรุปขั้นตอนการทำงาน

- สร้าง SUFFIX ARRAY เพื่อเรียงลำดับ SUFFIX ของสตริง
- สร้าง LCP ARRAY เพื่อหาความยาว PREFIX ที่เหมือนกันระหว่าง SUFFIXES ที่เรียงแล้ว
- หาค่าที่มากที่สุดใน LCP ARRAY เพื่อระบุ LONGEST REPEATED SUBSTRING

กระบวนการนี้ใช้เวลา $O(N \log N)$ สำหรับ SUFFIX ARRAY และ $O(N)$ สำหรับ LCP ARRAY ซึ่งรวมแล้วมีประสิทธิภาพสูงในการประมวลผลสตริงใหญ่ ๆ

A decorative graphic featuring a central light blue rectangle with a dark blue border. The Thai text 'ขอบคุณครับ' is centered within this rectangle. Surrounding the rectangle are several blue squares of various sizes, some of which are partially cut off by the edges of the frame.

ขอบคุณครับ

A decorative graphic featuring a central light blue rectangle with a dark blue border. The rectangle is surrounded by various sized blue squares of different shades, some of which are arranged in a pattern that suggests a stylized 'X' or a cross. The background is white.

ขอบคุณครับ