

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของงานวิจัย

การจัดเรียงข้อมูล (Sorting) เป็นวิธีสลับตำแหน่งของข้อมูลให้อยู่ในลำดับที่ถูกต้อง ไม่ว่าจะเรียงจากค่าน้อยไปหาค่ามากหรือจากค่ามากไปหาค่าน้อย การจัดเรียงข้อมูลนอกจากจะทำให้ข้อมูลมีลำดับถูกต้อง ประโยชน์อีกอย่างของการจัดเรียงข้อมูล คือ การนำมาใช้ช่วยค้นหาคีย์ที่ใช้การค้นแบบทวิภาค (Binary search) พื้นฐานการจัดเรียงข้อมูลเป็นงานหนึ่งที่สำคัญซึ่งเป็นหัวข้อที่ได้รับความสนใจศึกษาค้นคว้าจากนักวิชาการทางคณิตศาสตร์และคอมพิวเตอร์อย่างต่อเนื่องตั้งแต่อดีตจนถึงปัจจุบัน และเป็นวิชาพื้นฐานของการศึกษาด้านคอมพิวเตอร์

ในปัจจุบันขั้นตอนวิธีสำหรับแก้ปัญหาการจัดเรียงข้อมูลมีการค้นคิดขึ้นมามากมายหลายวิธี ซึ่งการจัดเรียงข้อมูลทางคอมพิวเตอร์แบ่งเป็นประเภทหลักๆ ได้ 2 ประเภท คือ การจัดเรียงภายในและการจัดเรียงภายนอก

การจัดเรียงภายใน (Internal sorting) คือ การจัดเรียงข้อมูลที่ข้อมูลทั้งหมดถูกเก็บไว้ในหน่วยความจำหลัก (RAM) การจัดเรียงแบบนี้ได้แก่ การจัดเรียงแบบบับเบิล (Bubble sort), การจัดเรียงแบบแทรก (Insertion sort), การจัดเรียงแบบเลือก (Selection sort), การจัดเรียงแบบเชลล์ (Shell sort), การจัดเรียงแบบเรดิคซ์ (Radix sort), การจัดเรียงแบบเร็ว (Quick sort) เป็นต้น

การจัดเรียงภายนอก (External sorting) คือ การจัดเรียงข้อมูลที่ใช้หน่วยความจำสำรองร่วมกับหน่วยความจำหลัก กรณีที่ข้อมูลมีจำนวนมากไม่สามารถเก็บไว้ในหน่วยความจำหลักเพื่อประมวลผลได้ทั้งหมด จึงต้องใช้หน่วยความจำสำรอง เช่น ฮาร์ดดิสก์ มาช่วยเก็บข้อมูลและแบ่งข้อมูลออกเป็นส่วนย่อยๆ แล้วทยอยนำเข้าไปจัดเรียงข้อมูลในหน่วยความจำหลัก การจัดเรียงแบบนี้ได้แก่ การจัดเรียงแบบผสาน (Merge sort)

งานวิจัยนี้จะขอกกล่าวถึงเฉพาะการจัดเรียงภายในเพราะว่าปัจจุบันการจัดเรียงภายในถูกใช้ในโปรแกรมประยุกต์ต่างๆ เป็นจำนวนมาก อีกทั้งเทคโนโลยีทางด้านฮาร์ดแวร์มีหน่วยความจำหลักขนาดใหญ่ทำให้คอมพิวเตอร์ทำงานกับข้อมูลปริมาณมากได้ และการจัดเรียงแบบภายในเป็นวิชาพื้นฐานของการศึกษาด้านคอมพิวเตอร์ วิธีการจัดเรียงภายในหลายๆ วิธี ต่างมีข้อดี ข้อเสีย และแนวทางการปรับปรุงประสิทธิภาพที่แตกต่างกัน การจัดเรียงภายในที่มี

ประสิทธิภาพการจัดเรียงที่ดีที่สุด [\[http://linux.wku.edu/~lamonml/algor/sort/sort.html\]](http://linux.wku.edu/~lamonml/algor/sort/sort.html) คือการจัดเรียงแบบเร็ว ต่อไปในงานวิจัยนี้ขอใช้คำว่า ควิกซอร์ต แทนคำว่า การจัดเรียงแบบเร็ว

ควิกซอร์ตเป็นวิธีการจัดเรียงภายในที่ได้รับการยอมรับว่ามีประสิทธิภาพในการจัดเรียงข้อมูลที่ดีที่สุด ควิกซอร์ตใช้เทคนิคการทำงานแบบแบ่งแยกและเอาชนะ (Divide and conquer) เวลาทำงานเฉลี่ยเท่ากับ $\Theta(n \log n)$ ทำงานได้เร็วกับข้อมูลขนาดใหญ่ และทำงานได้โดยใช้หน่วยความจำจำกัด

ควิกซอร์ตมีข้อด้อยที่สำคัญ คือ เมื่อนำข้อมูลที่เรียงลำดับมาทำการจัดเรียงพร้อมกับใช้การเลือกข้อมูลลำดับแรกเป็นตัวหลัก (Pivot) เวลาทำงานของควิกซอร์ตเท่ากับ $\Theta(n^2)$ จากกรณีดังกล่าวทำให้เกิดแนวคิดเพื่อปรับปรุงประสิทธิภาพและแก้ปัญหาคำกล่าวของควิกซอร์ตในหลายๆ แนวทางได้แก่

1.) ปรับปรุงการเลือกตัวหลักที่ใช้แบ่งกลุ่ม แนวทางนี้เสนอเทคนิคการเลือกข้อมูลแบบต่างๆ เช่น การสุ่มตำแหน่งของตัวหลัก (Randomize), เลือกข้อมูลตำแหน่งกลางของชุดข้อมูล (Middle element), การใช้ค่ากลางของชุดข้อมูลจากข้อมูล 3 ตำแหน่ง (Median of three), การใช้ค่ากลางของชุดข้อมูลจากวิธีการ Greater than less than เป็นต้น

2.) ปรับปรุงโดยใช้การจัดเรียงข้อมูลแบบอื่นแทนการเรียงด้วยควิกซอร์ตเมื่อข้อมูลย่อยมีขนาดเล็ก แนวทางนี้เสนอการศึกษาขนาดข้อมูลที่ดีและใช้เทคนิคการเปลี่ยนวิธีการจัดเรียงข้อมูลเป็นวิธีอื่นที่ง่ายกว่าและทำงานได้เร็วกับชุดข้อมูลขนาดเล็ก เช่น การจัดเรียงแบบแทรกหรือการจัดเรียงแบบเลือก แทนควิกซอร์ตเมื่อชุดข้อมูลย่อย (Subfiles) มีขนาดเล็กกว่าค่าที่กำหนด เช่น 10, 15, 25 เป็นต้น

3.) ปรับปรุงควิกซอร์ตด้วยเทคนิคอื่นๆ แนวทางนี้เสนอการใช้เทคนิคอื่นๆ เพื่อช่วยให้ควิกซอร์ตทำงานได้ดีขึ้น เช่น ใช้การตรวจสอบการเรียงลำดับของข้อมูลโดยคำสั่งการเปรียบเทียบคู่ลำดับติดที่อยู่กันหรือใช้การสลับคู่ลำดับที่เรียงไม่ถูกต้องเพื่อช่วยให้การเรียงลำดับเร็วขึ้น หรือใช้เทคนิคการแบ่งข้อมูลเป็นสองชุดตามลักษณะการเรียงตัวของคู่ลำดับ โดยแยกเป็นชุดของคู่ลำดับที่เรียงกับชุดคู่ลำดับที่ไม่เรียง

งานวิจัยนี้เป็นการปรับปรุงควิกซอร์ต โดยใช้เทคนิคผลต่างสืบเนื่องช่วยปรับปรุงประสิทธิภาพการทำงานของควิกซอร์ต โดยใช้สมบัติของผลต่างสืบเนื่อง (Successive difference) เพื่อช่วยการตรวจสอบการเรียงลำดับข้อมูลและทำการหยุดการจัดเรียงเมื่อชุดข้อมูลมีลำดับถูกต้องโดยไม่ต้องทำการเรียกซ้ำไปจนกระทั่งแบ่งกลุ่มย่อยข้อมูลไม่ได้ ตามเงื่อนไขของควิกซอร์ต

ทั่วไป ผู้วิจัยเรียกวิธีการนี้ว่าควิกซอร์ตที่ใช้ผลต่างสืบเนื่อง อีกทั้งควิกซอร์ตที่ใช้ผลต่างสืบเนื่องยังช่วยแก้ปัญหากรณีแย่งที่สุด ซึ่งเกิดจากการนำชุดข้อมูลที่เรียงลำดับแล้วมาจัดเรียง

ผลต่างสืบเนื่อง คือ ค่าของผลรวมทั้งหมดของผลต่างแบบไม่คิดเครื่องหมายของคู่ลำดับที่ติดกันในชุดข้อมูลจำนวน ค่าผลต่างสืบเนื่องของชุดข้อมูลที่เรียงลำดับ ทั้งเรียงลำดับจากค่ามากไปหาค่าน้อยหรือเรียงลำดับจากค่าน้อยไปหาค่ามากจะมีค่าเท่ากับผลต่างระหว่างค่าที่สูงที่สุดกับค่าที่ต่ำที่สุดในชุดข้อมูล ดังนั้นเมื่อพิจารณาในชุดข้อมูลที่เรียงลำดับจากค่าน้อยไปหาค่ามากแล้ว ผลต่างสืบเนื่องของชุดข้อมูล เท่ากับ ผลลบระหว่างค่าที่สูงที่สุด (ข้อมูลตัวสุดท้าย) กับค่าที่ต่ำที่สุด (ข้อมูลตัวแรก) ตัวอย่างชุดข้อมูลขนาด 6 ข้อมูล มีข้อมูลดังนี้ 9, 5, 1, 3, 8, 2

ข้อมูลค่ามากที่สุด คือ 9 ข้อมูลค่าน้อยที่สุด คือ 1 ผลต่างระหว่างค่ามากที่สุดกับค่าน้อยที่สุด คือ $9 - 1 = 8$

ผลต่างสืบเนื่องของชุดข้อมูลเท่ากับ $|9-5|+|5-1|+|1-3|+|3-8|+|8-2| = 4+4+2+5+6 = 21$

ผลต่างสืบเนื่องของชุดข้อมูลที่เรียงลำดับจากค่าน้อยไปหาค่ามาก เท่ากับ $|1-2|+|2-3|+|3-5|+|5-8|+|8-9| = 1+1+2+3+1 = 8$

ผลต่างสืบเนื่องของชุดข้อมูลที่เรียงลำดับจากค่ามากไปหาค่าน้อย เท่ากับ $|9-8|+|8-5|+|5-3|+|3-2|+|2-1| = 1+3+2+1+1 = 8$

ควิกซอร์ตที่ใช้ผลต่างสืบเนื่อง เป็นการใช้ผลต่างสืบเนื่องในการตรวจสอบการเรียงลำดับของชุดข้อมูล เมื่อพบว่าชุดข้อมูลมีลำดับถูกต้อง ควิกซอร์ตที่ใช้ผลต่างสืบเนื่องจะหยุดการทำงานทันที วิธีการนี้จะมีแนวคิดคล้ายกับวิธีการที่ใช้คำสั่งการเปรียบเทียบ [13, 14] ที่ใช้การตรวจสอบการเรียงลำดับของชุดข้อมูลแต่วิธีการผลต่างสืบเนื่องเป็นวิธีการที่ใช้งานได้กว้างกว่าเพราะสามารถนำวิธีผลต่างสืบเนื่องไปใช้ได้กับข้อมูลจำนวนทุกประเภทและใช้กับการจัดเรียงแบบอื่นๆ โดยขึ้นอยู่กับลักษณะการทำงานของวิธีการจัดเรียงนั้นๆ

ควิกซอร์ตที่ใช้ผลต่างสืบเนื่องกับปัญหากรณีแย่งที่สุดเมื่อนำข้อมูลเรียงลำดับมาจัดเรียงจากสมบัติของผลต่างสืบเนื่อง ในรอบแรกของการทำงานผลต่างสืบเนื่องจะตรวจสอบการเรียงลำดับข้อมูลและเมื่อพบว่าข้อมูลเรียงลำดับถูกต้องก็ให้หยุดการทำงานได้ทันทีโดยไม่ต้องจัดเรียงต่อไป ดังนั้นควิกซอร์ตที่ใช้ผลต่างสืบเนื่องจึงมีประสิทธิภาพของการทำงานสูงกับข้อมูลที่เรียงลำดับ เพราะเพียงรอบแรกของการทำงานวิธีการผลต่างสืบเนื่องจะช่วยหยุดการทำงานของควิกซอร์ตทันที

1.2 วัตถุประสงค์ของงานวิจัย

งานวิจัยนี้ศึกษาในหัวข้อเรื่องควิกซอร์ตที่ใช้ผลต่างสืบเนื่อง (A Successive Difference Quick sort) มีวัตถุประสงค์ดังนี้

1.2.1 ศึกษาถึงการนำวิธีการผลต่างสืบเนื่องมาประยุกต์ใช้กับควิกซอร์ต เพื่อการปรับปรุงประสิทธิภาพการทำงานให้กับควิกซอร์ต

1.2.2 ศึกษาถึงความเหมาะสม ข้อดี และข้อเสีย ของการนำวิธีผลต่างสืบเนื่องมาประยุกต์ใช้กับควิกซอร์ต

1.2.3 ศึกษาการใช้ควิกซอร์ตที่ใช้ผลต่างสืบเนื่องกับข้อมูลจำนวนเต็มแบบต่างๆ ที่มีการกำหนดลักษณะของความแตกต่างของค่าข้อมูล เพื่อหาลักษณะข้อมูลที่เหมาะสมกับ การนำควิกซอร์ตที่ใช้ผลต่างสืบเนื่องไปใช้งาน

1.2.4 การลดปัญหากรณีแย่งที่สุดเมื่อนำข้อมูลที่เรียงลำดับมาจัดเรียงด้วยควิกซอร์ตที่ใช้ผลต่างสืบเนื่อง

1.3 ขอบเขตของงานวิจัย

จากวัตถุประสงค์ที่กล่าวไว้ข้างต้น ในงานวิจัยนี้เป็นศึกษาวิธีการในการปรับปรุงการทำงานของการจัดเรียงแบบควิกซอร์ตด้วยการประยุกต์ใช้วิธีผลต่างสืบเนื่อง โดยมีขอบเขตของงานวิจัย ดังต่อไปนี้

1.3.1 ศึกษาการนำวิธีผลต่างสืบเนื่องมาประยุกต์ใช้กับควิกซอร์ต ในการวิจัยนี้จะทดลองกับข้อมูลที่เป็นจำนวนเต็ม เพื่อความสะดวกในการปรับปรุงประสิทธิภาพของขั้นตอนการหาผลต่างสืบเนื่อง และพิจารณาเฉพาะการเรียงแบบค่าน้อยไปค่ามากเท่านั้น

1.3.2 ปรับปรุงการทำงานของควิกซอร์ตด้วยวิธีการผลต่างสืบเนื่องด้วยภาษาจาวา ที่นำไปใช้ได้กับแพลตฟอร์มมากมาย

1.3.3 เปรียบเทียบการทำงานระหว่างควิกซอร์ตกับควิกซอร์ตด้วยผลต่างสืบเนื่องโดยวัดจากเวลาการทำงานและจำนวนการเรียกซ้ำ

1.4 ผลที่ได้รับ

- 1.4.1 คิวซอร์ตที่ใช้ผลต่างสืบเนื่อง ที่มีสมบัติดังนี้
- 1.4.1.1 หยุดทำงานเมื่อชุดข้อมูลมีการเรียงลำดับ
 - 1.4.1.2 ทำงานเร็วขึ้นในข้อมูลที่มีลักษณะการซ้ำกัน เรียงลำดับหรือเกือบเรียงลำดับ
 - 1.4.1.3 ใช้เวลาการทำงานเป็น $\Theta(n)$ ในข้อมูลที่เรียงลำดับ
 - 1.4.1.4 ใช้จำนวนการเรียกซ้ำลดลง
- 1.4.2 ผลการทดลองเปรียบเทียบระหว่างคิวซอร์ตกับคิวซอร์ตที่ใช้ผลต่างสืบเนื่องกับข้อมูลหลายๆ แบบ

1.5 รายละเอียดของวิทยานิพนธ์

รายละเอียดในส่วนอื่นๆ ที่เหลือของวิทยานิพนธ์ แบ่งออกเป็นส่วนต่างๆ ซึ่งสามารถอธิบายได้ดังหัวข้อต่อไปนี้

บทที่ 2. ทฤษฎี ผลงานวิจัยและงานเขียนอื่นๆ ที่เกี่ยวข้อง อธิบายถึงทฤษฎีต่างๆ ที่เกี่ยวข้องกับการวิจัยและเทคนิควิธีการต่างๆ ที่ใช้ในการปรับปรุงประสิทธิภาพของงานวิจัยที่เกี่ยวข้องกับคิวซอร์ต

บทที่ 3. วิธีการดำเนินงานวิจัย อธิบายรายละเอียดขององค์ประกอบที่ใช้ในงานวิจัยและขั้นตอนการศึกษาออกแบบคิวซอร์ตที่ใช้ผลต่างสืบเนื่อง

บทที่ 4. ผลการทดลอง อธิบายถึงรายละเอียดของผลการทดลองในแต่ละขั้นตอน

บทที่ 5. สรุปผลการศึกษาและข้อเสนอแนะ