



บทนำ

1.1 ความเป็นมาและสภาพของปัญหาในปัจจุบัน

ความต้องการใช้ภาษาไทยกับเครื่องคอมพิวเตอร์ ในปัจจุบันมีสูงมาก แต่เนื่องจากระบบคอมพิวเตอร์ถูกสร้างให้ใช้กับภาษาอังกฤษ ซึ่งตัวอักษรทุกตัวยบรรทัดเดียวกันต่างกับภาษาไทย ที่มีโครงสร้างเป็นหลายระดับ และไม่มีช่องว่างระหว่างคำ เพื่อแสดงความสั่นของคำ อย่างไรก็ตาม บรรดาผู้เชี่ยวชาญของกับคอมพิวเตอร์ของไทยได้พยายามต่อเติม ให้คอมพิวเตอร์รับภาษาไทยได้ โดยกำหนดรหัสภาษาไทยขึ้น ในส่วนที่ยื่นนอกเหนือรหัสภาษาอังกฤษ คือ ตั้งแต่รหัสแอสกี (ASCII) ที่ 128 ขึ้นไป จนได้เป็นรหัสภาษาไทยในมาตรฐานต่าง ๆ เช่น สมอ. เกษตร เป็นต้น

ในช่วงแรกของการใช้ภาษาไทยเหล่านี้ แม้จะใช้งานได้ และสามารถทำให้ผู้ใช้พอใจในระดับหนึ่ง แต่ก็ยังมีข้อบกพร่องอยู่ ทั้งเนื่องจากมาตรฐานของแต่ละรหัสไม่ตรงกัน และจากโครงสร้างของภาษาไทยเอง ทำให้เกิดปัญหา ทั้งในการจัดบรรทัดสำหรับแสดงผล และการพิมพ์รายงาน ที่ต้องการการเรียงลำดับที่ถูกต้อง ซึ่งข้อบกพร่องเหล่านี้จำเป็นต้องยอมรับ เนื่องจากสภาพการพัฒนาทางเทคโนโลยีในขณะนั้น ในปัจจุบัน เทคโนโลยีทางด้านภาษาไทยได้พัฒนาขึ้นมาก ข้อบกพร่องเหล่านี้จึงได้รับการแก้ไข เช่น ปัญหาเรื่องมาตรฐานของรหัสภาษาไทยได้มีการกำหนดให้เป็นมาตรฐานเดียว โดยใช้รหัสของสมอ. เป็นหลัก ส่วนปัญหาเรื่องการจัดบรรทัดสำหรับการแสดงผล เวลานี้ไม่เป็นปัญหาอีกต่อไป เพราะได้มีการผลิตแผ่นวงจรสำหรับการจัดบรรทัดภาษาไทย โดยเฉพาะ ดังนั้น จะเห็นได้ว่า ข้อบกพร่องในการใช้ภาษาไทยกับเครื่องคอมพิวเตอร์ กำลังได้รับการแก้ไขให้ลดน้อยลงเรื่อย ๆ แต่ยังมีเหลืออยู่ ซึ่งหนึ่งในข้อบกพร่องที่เหลืออยู่นี้และเป็นปัญหาอย่างมากสำหรับผู้ใช้ในปัจจุบัน นั่นคือ การจัดเรียงข้อมูลภาษาไทยให้ตรงตามพจนานุกรมเพื่อเก็บไว้ในระบบคอมพิวเตอร์ หรือพิมพ์เป็นรายงานให้เรียงลำดับได้

อนึ่ง วิทยานิพนธ์เกี่ยวกับการจัดเรียงลำดับนี้ ได้มีนิตยสารปริทัศน์ของภาควิชาวิทยาศาสตร์คอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ได้เคยทำเป็นวิทยา-

แม่เข้ามาแล้ว ดั่ง!

(1) วิทยานิพนธ์เรื่องการศึกษาโครงสร้างโปรแกรมสำเร็จรูป สำหรับการจัดรวมข้อมูลอย่างมีลำดับ โดยนายเลิศชาย ฤกษ์ศรีหิ ปริญญาการศึกษา 2524 ซึ่งได้ศึกษาเกี่ยวกับโครงสร้างโปรแกรมและการใช้งานของโปรแกรมสำเร็จรูป สำหรับการจัดรวมข้อมูล อย่างมีลำดับ หมายเลข 5746 - เอสเอ็ม1 สำหรับเครื่องคอมพิวเตอร์ ไอบีเอ็ม 370/138 ซึ่งทำงานภายใต้การควบคุมของระบบ ดอส/วีเอส

(2) วิทยานิพนธ์เรื่องการพัฒนาคำสั่งการเรียงลำดับข้อมูล สำหรับอินเตอร์พรีตเตอร์ภาษาเบสิก โดยนางสาวจตุรนต์ เรืองหทัยธรรม ปริญญาการศึกษา 2528 ซึ่งได้ศึกษาขั้นตอนการทำงาน ของอินเตอร์พรีตเตอร์ภาษาเบสิก (Microsoft Basic-80) ของบริษัทไมโครซอฟท์ สำหรับไมโครคอมพิวเตอร์ รุ่นแอปเปิล ๒ พลัส (Apple II Plus) ที่มีขนาดหน่วยความจำหลัก 64 กิโลไบต์ ทำงานภายใต้ระบบดำเนินการซีพี/เอ็ม (CP/M) รวมทั้งศึกษาการเรียงลำดับข้อมูลภาษาอังกฤษและตัวเลข ที่เก็บเป็นรหัสแอสกี

1.1.1 การเรียงลำดับข้อมูล (Sorting)

การจัดเรียงลำดับข้อมูลด้วยเครื่องคอมพิวเตอร์ หมายถึง วิธีการจัดเรียงลำดับข้อมูลตามลำดับของคีย์ (Key) อาจจะเรียงจากน้อยไปมาก (Ascending) หรือจากมากไปน้อย (Decending) โดยเรียงแบบต่อเนื่องกัน ข้อมูลนั้นอาจอยู่ในสื่อกลาง (Media) เช่น แผ่นจานแม่เหล็ก (Diskette) หรือจะอยู่ภายในหน่วยความจำของเครื่องคอมพิวเตอร์ ในการจัดเรียงลำดับข้อมูลนั้น จำเป็นต้องใช้หน่วยความจำในการเรียงลำดับ ซึ่งอาจเป็นหน่วยความจำหลัก (Main Memory) หรือหน่วยความจำสำรอง (Auxiliary Memory) ขึ้นอยู่กับวิธีการเรียงลำดับข้อมูล ซึ่งแบ่งออกเป็น 2 ลักษณะ ดังนี้

1.1.1.1 การเรียงลำดับแบบภายใน (Internal Sorting) เป็นการ

เรียงลำดับโดยใช้เฉพาะหน่วยความจำหลักของเครื่อง ระเบียบทั้งหมดจะเข้ามาอยู่ในหน่วยความจำหลักพร้อม ๆ กัน มีความเร็วสูง มีขั้นตอนวิธี (Algorithm) มากมายขึ้นอยู่กับความเหมาะสม รายละเอียดของข้อมูล และลักษณะการเกิดของข้อมูล มีวิธีสำคัญ 3 วิธี ดังนี้

1) การเรียงลำดับแบบฟอง (Bubble Sort)

เป็นวิธีที่ง่ายที่สุด ใช้หลักการนำคีย์ตัวหนึ่งมาเปรียบเทียบกับคีย์ที่อยู่ถัดไป ทำการสลับที่ระหว่างระเบียบที่มากกว่ากับระเบียบที่น้อยกว่า ดังนั้น ระเบียบที่มากกว่าที่สุด จะค่อย ๆ เลื่อนไปอยู่ในตำแหน่งท้าย ทำเช่นนี้เรื่อยไป จนกว่าทุกระเบียบจะอยู่ใน

ลำดับที่ถูกต้อง

2) การเรียงลำดับแบบควิกซอร์ต (Quicksort)

เป็นวิธีที่นิยมกันแพร่หลายที่สุดวิธีหนึ่ง และจะเป็นวิธีที่มีประสิทธิภาพที่สุดของการเรียงลำดับแบบภายใน ซึ่งใช้ได้กับโปรแกรมปกติทั่วไป วิธีการคล้ายกับวิธีแรก กล่าวคือ ครั้งแรกจะหาระเบียง 1 ระเบียน เคลื่อนไปหาตำแหน่งที่ควรจะอยู่ทางขวามือ ในขณะที่หาตำแหน่งทางขวามือ ก็จัดระเบียบทางด้านซ้ายใหม่ค่าน้อยกว่าหรือมากกว่าหรือเท่ากับตำแหน่งที่หาอยู่ และระเบียบทางขวาซึ่งมีค่ามากกว่าหรือเท่ากับนั้นคือ เพิ่มข้อมูลจะถูกแบ่งออกเป็นเพิ่มข้อมูลย่อย ๆ และทำวิธีการเช่นนี้ จะได้ข้อมูลที่เรียงลำดับ

3) การเรียงลำดับแบบเชลล์ (Shellsort)

เป็นการเรียงลำดับที่ใช้การแทรกแบบเชิงเส้น (Linear Insertion Sort) ซ้ำ ๆ กัน และเป็นที่ยอมรับมากที่สุด เริ่มจากการเปรียบเทียบข้ามตำแหน่ง โดยการแบ่งระเบียบออกเป็นกลุ่ม นำมาเรียงลำดับแต่ละกลุ่มก่อน แล้วแบ่งเป็นกลุ่มอีกให้เล็กลง และเรียงลำดับในแต่ละกลุ่มอีกทำเช่นนี้จนกระทั่งเหลือกลุ่มเดียว

1.1.1.2 การเรียงลำดับแบบภายนอก (External Sorting)

เทคนิคต่างจากการจัดเรียงลำดับแบบภายใน เพราะว่าต้องใช้โครงสร้างข้อมูลมาเปรียบเทียบ และแน่นอนว่ามาก ในเรื่องเวลาดำเนินการของข้อมูลเข้า (Input) หรือข้อมูลออก (Output) รวมทั้งต้องเรียนรู้เกี่ยวกับโครงสร้างข้อมูล, การติดต่อ หรือการจัดการหน่วยความจำภายนอกอย่างไร และเลือกใช้หรือวิเคราะห์ขั้นตอนวิธีอย่างไร ตัวอย่างเช่น

1) การเรียงลำดับผสานได้ดล (Balance Merge Sort)

เป็นวิธีการจัดเรียงลำดับแบบภายนอก ท้ายที่สุด ซึ่งจะแบ่งเพิ่มข้อมูลออกเป็น 2 ส่วน ในหมีจำนวนระเบียบในใกล้เคียงกัน แล้วแต่ละกลุ่มก็จัดเรียงลำดับเมื่อแยกการจัดเรียงลำดับแล้ว ต่อมา กรวม 2 กลุ่มเข้าด้วยกันเป็นกลุ่มเดียวกัน

2) การเรียงลำดับควิกซอร์ตภายนอก (External Quick-

sort)

วิธีนี้เป็นวิธีที่สมบรูณ์สำหรับการจัดเรียงลำดับแบบภายนอกพื้นฐานของขั้นตอนวิธีนี้ เหมือนกับการเรียงลำดับควิกซอร์ตภายใน

1.1.2 ขั้นตอนก่อน/หลังการเรียงลำดับ (Pre/Post Sort Phase)

ในการเรียงลำดับข้อมูลภาษาไทย จะกระทบกับข้อมูลโดยตรงไม่ได้ เพราะผลลัพธ์จะไม่ตรงตามหลักภาษาไทย สาเหตุมาจากการวางตำแหน่ง และคาร์ริสแอสกีของสระและวรรณยุกต์ ดังนั้น ก่อนที่จะทำการเรียงข้อมูลที่เป็นภาษาไทย จะต้องผ่านขั้นตอนก่อนการเรียงลำดับ (Pre-Sort Phase) สำหรับเปลี่ยนรูปข้อมูลภาษาไทยให้เหมาะสม เพื่อว่าเมื่อส่งข้อมูลผ่านขั้นตอนการเรียงลำดับแล้ว จะได้ผลการเรียงที่ถูกต้อง

การดำเนินการต่าง ๆ ในขั้นตอนก่อนการเรียงลำดับ ประกอบด้วยวิธีการ ดังต่อไปนี้

1.1.2.1 การสลับตำแหน่ง

สระหน้าหน้าพยัญชนะ เช่น เ, แ, โ, ใ, ไ จะถูกสลับตำแหน่งกับพยัญชนะ เช่น

เวลา... เปลี่ยนเป็น วเวลา...

สำหรับวรรณยุกต์กระทำในตัวเองเดียวกัน เช่น

พอ... เปลี่ยนเป็น พอ...

1.1.2.2 การเพิ่มหรือลดน้ำหนักของอักขระ

การกำหนดรหัสแอสกีของอักขระไทยบางตัว ส่งผลให้การเรียงลำดับไม่ถูกต้อง จึงต้องกำหนดค่าให้อักขระนั้นใหม่ เป็นค่าน้ำหนัก (Weight) ที่ใช้ในการเรียงเท่านั้น ไม่ใช่รหัสแอสกีใหม่ของอักขระนั้น ซึ่งการกำหนดค่าน้ำหนักอาจกำหนดให้เพิ่มขึ้นหรือลดลงก็ได้ เพื่อให้ได้ผลการเรียงที่ถูกต้อง และเปลี่ยนค่าน้ำหนักคืน ภายหลังจากการเรียงลำดับ

1.1.2.3 การยบรูป

การเรียงลำดับภาษาไทย ตามข้อกำหนดที่ปรากฏในพจนานุกรมไทย ฉบับราชบัณฑิตยสถาน พ.ศ. 2525 กำหนดให้เรียงลำดับ ดังนี้

ก-ฮ,

ะ, ั, ็, า, ำ, ิ, ี, ึ, ุ, ู, เ, แ, โ, ใ, ไ, เ้, ็้, ำ้, ิ้, ี้, ึ้, ุ้, ู้, เ๊, แ๊, โ๊, ใ๊, ไ๊

ดังนั้น จากกฎเกณฑ์ดังกล่าว จะช่วยให้การเรียงลำดับมีความถูกต้องมากขึ้น โดยอาศัยวิธีการยบรูปกลุ่มของสระและวรรณยุกต์เหล่านี้ เป็นอักขระพิเศษ 1 ตัว ซึ่งมีน้ำหนักที่เหมาะสม เช่น

เกาะ... เปลี่ยนเป็น กX...

โดย X เป็นอักขระที่มีน้ำหนักเฉพาะ

ในขั้นตอนก่อนการเรียงลำดับ จะอาศัยวิธีการทั้ง 3 แบบร่วมกัน เพื่อเปลี่ยนข้อมูลเดิม ในอยู่ในรูปที่สามารถใช้โปรแกรมเรียงลำดับ ทำการเรียงตามปกติ แต่ได้ผลถูกต้องตามพจนานุกรม จากนั้น จึงผ่านขั้นตอนหลังการเรียงลำดับ (Post-Sort Phase) เพื่อดึงเอารหัสแอสกีเดิมมาใช้

1.2 ขอบเขตการวิจัย

การทำวิจัยจะเป็นการสร้างโปรแกรมเรียงลำดับข้อมูลภาษาไทยขึ้น ในอยู่ในรูปของโปรแกรมย่อย โดยมีรายละเอียดดังนี้

- 1) โปรแกรมหลักสามารถเรียกใช้โปรแกรมเรียงลำดับข้อมูล ได้จากคำสั่งภายในตัวโปรแกรมหลักเอง ไม่จำเป็นต้องกลับไปเข้าสู่โปรแกรมควบคุมระบบก่อน
- 2) โปรแกรมหลัก ที่เขียนด้วยภาษาเบสิก (BASIC), ฟอรัแทรน (FORTRAN), ปาสกาล (Pascal) และซี (C) ของบริษัทไมโครซอฟท์ (Microsoft) สามารถเชื่อมโยง (link) เข้ากับโปรแกรมการเรียงลำดับข้อมูลได้
- 3) โปรแกรมเรียงลำดับข้อมูล สามารถเรียงได้ทั้งตัวเลขและตัวอักษร ทั้งภาษาไทยและภาษาอังกฤษบนกัน
- 4) คีย์ที่ใช้เรียงสามารถมีได้ถึง 7 คีย์ โดยแต่ละคีย์สามารถกำหนดว่าเป็นคีย์ภาษาไทย หรือภาษาอังกฤษ และเรียงจากมากไปน้อย หรือจากน้อยไปมาก ได้โดยอิสระไม่ชนกัน
- 5) ในกรณีคีย์เป็นภาษาไทย จะยึดหลักการเรียงให้ใกล้เคียงกับพจนานุกรมภาษาไทยให้มากที่สุด
- 6) การเรียงลำดับทำแบบการเรียงภายนอก โดยใช้หน่วยความจำหรือเนื้อที่บนแผ่นงานแม่เหล็กทางทั้งหมด (Utilize Memory and Disk Space) เป็นเนื้อที่ในการเรียง
- 7) รหัสภาษาไทยที่ใช้ สามารถใช้รหัสของสมอ. และเกษตร
- 8) โปรแกรมเรียงลำดับข้อมูลเขียนด้วยภาษาซี ของบริษัทไมโครซอฟท์
- 9) เครื่องคอมพิวเตอร์ที่ใช้ จะเป็นไมโครคอมพิวเตอร์ ระดับ 16 บิตขึ้นไป มีเอ็มเอส-ดอส (MS-DOS) หรือ พีซี-ดอส (PC-DOS) เป็นโปรแกรมควบคุมระบบ
- 10) สามารถใช้เป็นโปรแกรมเรียงลำดับแบบลำพัง (Standalone) ได้



1.3 วัตถุประสงค์ในการวิจัย

- 1) ศึกษาคำสั่งของภาษาซี
- 2) ศึกษาขั้นตอนการทำงานของโปรแกรมควบคุมระบบ
- 3) วิเคราะห์และคัดเลือกทฤษฎี การเรียงลำดับข้อมูลแบบภายนอก ที่เหมาะสมและมีประสิทธิภาพ
- 4) ศึกษาความแตกต่างระหว่างรหัสภาษาไทยแต่ละรหัส
- 5) ศึกษาวิธีการและรูปแบบในการติดต่อ ระหว่างภาษาเบสิก, พอร์แทรน, และ ปาลกาล กับภาษาซี
- 6) เขียนและทดสอบโปรแกรมเรียงลำดับข้อมูล
- 7) จัดทำคู่มือการใช้

1.4 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

- 1) เพื่อเอื้ออำนวยให้ผู้ใช้ ได้มีโปรแกรมจัดเรียงลำดับข้อมูลภาษาไทยไว้ใช้ โดยรวมเข้าเป็นส่วนหนึ่งของ โปรแกรมประยุกต์
- 2) เพื่อให้ได้โปรแกรมเรียงลำดับข้อมูลภาษาไทยที่ถูกต้อง ตามพจนานุกรม
- 3) เพื่อให้การใช้ทรัพยากรของระบบเป็นไปอย่างมีประสิทธิภาพ ซึ่งส่งผลให้การ ทำงานรวดเร็วขึ้น
- 4) เพื่อช่วยให้อุปกรณ์ไม่ต้อง เสียเวลาในการพัฒนาโปรแกรมเรียงลำดับข้อมูลตนเอง
- 5) เพื่อให้ผู้ใช้ที่รู้รหัสภาษาไทยต่าง ๆ เช่น รหัสของสมอ. และ เกษตร สามารถเรียกใช้โปรแกรมเรียงลำดับข้อมูลภาษาไทยนี้ได้
- 6) เพื่อให้ผู้ใช้ที่เขียนโปรแกรมประยุกต์ด้วยภาษาระดับสูง เช่น เบสิก, พอร์แทรน, ปาลกาล และซี สามารถเรียกใช้โปรแกรมเรียงลำดับข้อมูลได้สะดวกและง่ายที่สุด
- 7) เป็นแนวทางในการพัฒนาการ เรียงลำดับข้อมูลภาษาไทย ให้มีประสิทธิภาพดียิ่งขึ้นในอนาคต