



บทที่ 6

สรุปผลการวิจัยและข้อเสนอแนะ

6.1 สรุปผลการวิจัย

6.1.1 วิธีการ THUFF

วิธีการ HUFF เป็นวิธีการอัดข้อมูลที่ลดขนาดของข้อมูลลง โดยการลดความเหลือเพื่อที่เกิดจากการใช้งานไม่เท่ากันของรหัสต่าง ๆ แต่วิธีการ THUFF จะสามารถลดขนาดของข้อมูลลงจากความเหลือเพื่อ 2 ลักษณะคือ ความเหลือเพื่อที่เกิดจากการใช้งานไม่เท่ากันของรหัสต่าง ๆ และความเหลือเพื่อที่เกิดจากรูปแบบที่ใช้งานมาก ดังนั้นวิธีการ THUFF จึงสามารถให้ประสิทธิภาพการอัดสูงกว่าวิธีการ HUFF มาก แต่ถ้าข้อมูลมีลักษณะความเหลือเพื่อในแบบอื่นรวมอยู่ด้วยเช่น การเกิดซ้ำ ๆ ของรหัสใดรหัสหนึ่ง หรือข้อมูลมีการใช้ภาษาอื่นรวมอยู่ด้วยก็จะมีผลกระทบต่อประสิทธิภาพการอัดข้อมูลด้วย การนำวิธีการ THUFF ไปใช้งานจะให้ประสิทธิภาพที่ดีมากกับข้อมูลภาษาไทยล้วน

ลักษณะรหัสของวิธีการ THUFF จำนวนรหัสที่ใช้ในการแทนรหัสจะคงที่ และมีรูปแบบรหัสคงที่ตลอด รหัสแบ่งออกเป็น 2 กลุ่ม รหัสกลุ่มแรกเป็นรหัสแอสกี 256 รหัส รหัสกลุ่มที่ 2 เป็นกลุ่มรหัสของรหัสคำไทยที่ใช้ ซึ่งจำนวนรหัสจะขึ้นอยู่กับขนาดของรหัสที่ใช้งาน ถ้าใช้ขนาดของรหัส 8 บิต รหัสกลุ่มนี้จะมีทั้งหมด 255 รหัส แต่ถ้าขนาดรหัสเป็น 9 บิต จำนวนรหัสในกลุ่มนี้จะมี 511 รหัส ดังนั้นวิธีการ THUFF จะมีรหัสใหม่จำนวนคงที่ และเป็นรหัสของคำไทยเท่านั้น จะไม่มีการเรียนรู้รหัสอื่น ๆ อีก

เวลาที่ใช้ในการอัดข้อมูลจะมากกว่าวิธีการเดิม เพราะจะต้องใช้เวลาในการตรวจสอบคำกับคำไทยในตารางในแต่ละรอบที่อ่านข้อมูล

การทำงานของวิธีการ THUFF เป็นการทำงานแบบ 2 รอบ ดังนั้นวิธีการนี้ไม่สามารถนำไปใช้งานในการอัดข้อมูลลักษณะเวลาจริงได้ (Real Time)

6.1.2 วิธีการ TLZW

วิธีการ TLZW ให้ประสิทธิภาพการอัดข้อมูลที่ดีกว่าวิธีการ LZW แม้ว่าจะใช้เวลาส่วนหนึ่งในการสร้างรหัสคำไทย แต่จำนวนครั้งของการแทนรหัสจะน้อยลง ดังนั้นเวลาที่ใช้ในการอัดข้อมูลและการขยายข้อมูลจึงใกล้เคียงกับวิธีการเดิม ในกรณีที่ข้อมูลมีขนาดไม่ใหญ่มาก

เวลาที่ใช้ในการอัดข้อมูลและการขยายข้อมูลอาจจะมากกว่าวิธีการเดิม แต่จะสามารถให้ประสิทธิภาพเพิ่มขึ้น

ลักษณะรหัสของวิธีการ TLZW จำนวนรหัสที่ใช้ในการแทนรหัสมีจำนวนคงที่ ซึ่งขึ้นอยู่กับขนาดสูงสุดของรหัสที่ใช้ ถ้าใช้ขนาดรหัส 12 บิต จำนวนรหัสจะเท่ากับ 4096 แต่ถ้าใช้ขนาดรหัส 13 บิต จำนวนรหัสจะเท่ากับ 8192 รหัส รหัสแบ่งออกเป็น 3 กลุ่ม รหัสกลุ่มแรก เป็นรหัสแอสกี 256 รหัส รหัสกลุ่มที่ 2 เป็นรหัสที่สร้างขึ้นจากการกระจายค่าไทย จำนวน 525 รหัส จากค่าไทย 255 คำ หรือจำนวน 1057 รหัสจากค่าไทย 511 คำ รหัสกลุ่มที่ 3 เป็นรหัสที่ได้จากการเรียนรู้จากลักษณะของข้อมูลที่ ในขณะที่ รหัส 2 กลุ่มแรกเป็นรหัสที่มีรูปแบบคงที่ตลอดไป แต่รหัสกลุ่มที่ 3 เป็นรหัสที่มีรูปแบบเปลี่ยนแปลงไปตามลักษณะของข้อมูล

วิธีการ TLZW เป็นวิธีการที่เหมาะสมสำหรับข้อมูลภาษาไทยทั้ง 2 ลักษณะ วิธีการทำงานเป็น 1 รอบ จึงสามารถนำไปใช้งานในลักษณะเวลาจริงได้

6.1.3 ข้อสรุปสำหรับวิธีที่เหมาะสมที่สุด

การเปรียบเทียบความเหมาะสมระหว่างวิธีการ THUFF กับวิธีการ TLZW สำหรับนำมาใช้ในการอัดข้อความภาษาไทย มีปัจจัยในการพิจารณาดังนี้

1. ประสิทธิภาพการอัด วิธีการที่ปรับเปลี่ยนสามารถให้ประสิทธิภาพการอัดที่ดีขึ้น และการเพิ่มจำนวนคำไทยมีผลทำให้ประสิทธิภาพการอัดเพิ่มขึ้น วิธีการ THUFF ให้ประสิทธิภาพการอัดกับข้อความภาษาไทยล้วนดีกว่าข้อความภาษาไทยทั่ว ๆ ไป ส่วนวิธีการ TLZW ให้ประสิทธิภาพการอัดกับข้อความภาษาไทยทั่ว ๆ ไปดีกว่าข้อความภาษาไทยล้วน
2. เวลาที่ใช้ในการอัดข้อมูล วิธีการ THUFF ใช้เวลาในการอัดข้อมูลเพิ่มขึ้นมาก ในขณะที่วิธีการ TLZW ใช้เวลาใกล้เคียงกับวิธีการเดิม
3. เวลาที่ใช้ในการขยายข้อมูล วิธีการที่ปรับเปลี่ยนทั้ง 2 วิธีใช้เวลาใกล้เคียงกับวิธีการเดิม
4. ความซับซ้อนของอัลกอริทึม วิธีการ THUFF มีอัลกอริทึมซับซ้อนมากกว่าวิธีการ TLZW เพราะต้องมีการตรวจสอบกลุ่มอักษรในข้อความกับคำไทย ในขั้นตอนการนับความถี่ และการเข้ารหัสข้อมูล ซึ่งเป็นขั้นตอนที่ยุ่งยากพอสมควรเนื่องจากลักษณะของภาษาไทยเองที่ไม่มีการใช้ช่องว่างระหว่างคำดังเช่นภาษาอื่น ๆ ทำให้ผลกับเวลาที่ใช้ในการอัดข้อมูล

จากปัจจัยดังกล่าวข้างต้นสามารถสรุปได้ว่า วิธีการ TLZW เหมาะสมสำหรับการนำไปใช้ในการอัดข้อความภาษาไทยมากกว่าวิธีการ THUFF

6.2 ข้อเสนอแนะ

เนื่องจากวิธีการ TLZW ที่พัฒนาขึ้นใช้ขนาดรหัสสูงสุด 12 บิต การนำไปใช้งาน อาจเพิ่มขนาดรหัสสูงสุดเป็น 13 บิต ซึ่งจะสามารถช่วยเพิ่มประสิทธิภาพการอัดข้อมูลได้อีกทางหนึ่ง

วิธีการอัดข้อมูลภาษาไทยนี้ สามารถนำไปประยุกต์ใช้งานร่วมกับโปรแกรมประมวลผล คำภาษาไทย (Thai Word Processing) หรืออาจนำไปพัฒนาเป็นโปรแกรมประยุกต์จัดเก็บ เอกสารภาษาไทย (Thai File Archives)



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย