



บทที่ 8

สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้ได้พัฒนาโปรแกรมพจนานุกรมอิเล็กทรอนิกส์ โดยสามารถสืบค้นเพิ่มเติม แก้ไขและลบข้อมูลในพจนานุกรมได้ และสามารถใช้งานร่วมกับจุฬารีก นอกจากนี้เนื่องจากพจนานุกรมใช้พื้นที่จัดเก็บค่อนข้างมาก จึงได้ทำการบีบอัดเพื่อลดขนาดข้อมูล

การบันทึกคำศัพท์และความหมายนั้น ได้กำหนดโครงสร้างข้อมูลไว้ 2 รูปแบบ คือ พจนานุกรมกำหนดโครงสร้างมาให้ และผู้ใช้สามารถกำหนดชนิดของข้อมูลที่ต้องการในภาคแสดงความหมายได้

การสืบค้นคำศัพท์ สามารถสืบค้นเฉพาะส่วนต้นของคำศัพท์ และสามารถคัดลอกข้อความจากโปรแกรมประยุกต์อื่นเพื่อนำมาค้นหาความหมาย หรือคัดลอกข้อความในภาคแสดงความหมายเพื่อนำไปแทรกในข้อความของโปรแกรมประยุกต์อื่น นอกจากนี้ยังสามารถกำหนดค่าในส่วนของคำเหมือนและคำตรงข้ามเพื่อใช้เป็นคำศัพท์สำหรับค้นหาความหมายต่อไปได้

งานวิจัยนี้ได้ใช้โครงสร้างการจัดเก็บและสืบค้นข้อมูลโดยวิธีดิเจิตอล เลิซ-ทรี (ดี-เอส ทรี) คือเป็นการใช้อักษรแต่ละตัวในคำศัพท์มาเปรียบเทียบทีละตัว โดยสร้างดี-เอส ทรี จากกลุ่มของคำศัพท์แล้วจึงทำการสืบค้นจาก ดี-เอส ทรี ที่สร้างขึ้นมานี้ และในการแสดงดี-เอส ทรี ในเชิงโปรแกรม จะใช้โครงสร้างข้อมูลที่เรียกว่าดับเบิลอะเรย์ทรี ซึ่งการสืบค้นโดยวิธีนี้จะใช้อักษรแต่ละตัวที่ประกอบเป็นคำศัพท์นั้นๆ มาใช้เป็นดัชนีเข้าไปค้นหาในตาราง ซึ่งจะได้ผลลัพธ์เป็นค่าฐานเริ่มต้นสำหรับตัวอักษรตัวต่อไป ดำเนินการค้นหาเรื่อยไปจนถึงตัวสุดท้าย จะได้ผลลัพธ์เป็นค่าบอกตำแหน่งของข้อมูลสำหรับคำศัพท์นั้นๆ ความเร็วในการค้นหาจึงขึ้นกับจำนวนตัวอักษรของคำศัพท์ โดยไม่ขึ้นกับจำนวนคำที่ประกอบขึ้นเป็นดับเบิลอะเรย์ทรี เพราะฉะนั้นเวลาที่ใช้ในการสืบค้นคำศัพท์เป็นไปตามฟังก์ชัน $O(k)$ โดยที่ k คือความยาวของคำศัพท์ที่ต้องการสืบค้น

การใช้งานร่วมกับจุฬารีก ในส่วนนี้จะใช้วิธีเรียกโปรแกรมจุฬารีกขึ้นมาทำงาน ขณะเดียวกันก็แทรกเมนูพจนานุกรมลงไปด้วย เมื่อมีการเรียกใช้เมนูนี้ โปรแกรมพจนานุกรมก็จะทำงาน ผู้ใช้สามารถป้อนคำศัพท์เพื่อค้นหาความหมาย คัดลอกคำในจุฬารีกมาค้นหาความหมาย หรือคัดลอกข้อความในภาคแสดงความหมายไปแทรกในข้อความของจุฬารีกได้ การทำงานในส่วนนี้ เนื่องจากเป็นการดึงเมสเสจจากคนละโปรแกรมประยุกต์จึงต้องใช้ช่องเกี่ยวโยง (Hooks) เพื่อช่วยในการดึงเมสเสจของทั้งระบบ เมื่อตรวจสอบพบว่าเป็นเมสเสจที่เกิดจากการเลือกเมนูพจนานุกรม โปรแกรมก็จะทำงาน ซึ่งจากการทดสอบโปรแกรม พบว่าสามารถทำงานได้ตามที่ต้องการ

การลดขนาดของข้อมูล ในงานวิจัยนี้ได้ใช้วิธีแอลแชนด์ดับบลิว ซึ่งวิธีนี้ในขณะที่ทำการลดขนาด และขยายข้อมูลจะสร้างรหัสไปด้วย ทำให้ไม่ต้องสิ้นเปลืองเนื้อที่ในการจัดเก็บตารางสำหรับการถอดรหัส และเนื่องจากการกำหนดรหัสด้วยจำนวนบิตมากทำให้สามารถกำหนดรหัสได้มากและประสิทธิภาพการ บีบอัดข้อมูลสูง แต่ในกรณีที่เพิ่มข้อมูลก่อนการบีบอัดมีขนาดเล็กอยู่แล้วก็จะทำให้ข้อมูลที่ผ่านการบีบอัดมีขนาดใหญ่ขึ้น ฉะนั้นในการวิจัยนี้จะใช้จำนวนบิตที่ 9 บิตในตอนเริ่มต้นและหากมีการกำหนดรหัสเกินจำนวน BUMP_CODE คือ 257 โปรแกรมจะเปลี่ยนไปใช้จำนวนบิต 10 บิต และเพิ่มเป็น 11, 12 จนกระทั่งสูงสุดที่ 15 บิต

จากการทดสอบปรากฏว่าสามารถลดขนาดข้อมูลพจนานุกรมลงได้ประมาณร้อยละ 60

ข้อเสนอแนะ

โครงสร้างการจัดเก็บคำศัพท์โดยวิธีดี-เอส หรือ นี้การดำเนินการกับคำศัพท์นั้น คำศัพท์ทั้งหมด จะต้องถูกจัดเก็บบนหน่วยความจำหลัก หากคำศัพท์มีจำนวนมากขึ้น หน่วยความจำหลักซึ่งมีจำกัดอาจ ไม่สามารถรองรับได้ ฉะนั้นหากมีการปรับปรุงการดำเนินการในส่วนนี้หรือมีวิธีการจัดการเกี่ยวกับหน่วย ความจำหลักที่ดีโดยที่ไม่ทำให้ประสิทธิภาพในการทำงานลดลงแล้ว จะทำให้พจนานุกรมสามารถรองรับ คำศัพท์ปริมาณมากๆ ได้

การปิดท้ายคำศัพท์ (pad) ด้วยตัวสิ้นสุดคำ (terminator) ซึ่งในที่นี้ใช้ # ก็เพื่อหลีกเลี่ยงปัญหา คำในคำ เช่น use กับ user จะพบว่ามี use เป็นคำในคำของ user ซึ่งการที่คำทุกคำถูกทำให้ลงท้ายด้วย ตัวสิ้นสุดคำ ก็จะไม่ทำให้เกิดปัญหาเช่นนี้ขึ้น ดังนั้นคำทุกคำจะไปสิ้นสุดที่ TAIL เหมือนกัน การปิดท้าย ด้วยตัวสิ้นสุดคำนี้ ทำให้เกิดโอเวอร์เฮด ดังนั้นหากมีการปรับปรุงอัลกอริทึมใหม่โดยเพิ่มส่วนของ โปรแกรมให้สามารถรองรับในส่วนนี้ จะทำให้ลดจำนวนลูป (loop) ในการสืบค้นลงไป (ส่วนของตัว #)

เนื่องจากข้อมูลในพจนานุกรมต้องการเนื้อที่จัดเก็บมาก ถึงแม้ในการวิจัยนี้จะได้ทำการลด ขนาดข้อมูลในขณะที่ไม่ได้ใช้พจนานุกรม และทำการขยายเมื่อเรียกใช้พจนานุกรม ผู้วิจัยขอเสนอให้ พัฒนาการขยายข้อมูลเฉพาะเรคคอร์ดที่เรียกใช้ ซึ่งจะช่วยให้ประหยัดเนื้อที่ในการจัดเก็บได้มาก

จุฬาลงกรณ์มหาวิทยาลัย