

บทสรุปและข้อเสนอแนะ

พจนานุกรมอิเล็กทรอนิกส์ภาษาไทยที่สร้างขึ้นนั้นสามารถนำมาใช้ในงานประมวลผลทางด้านภาษาศาสตร์ ซึ่งจำเป็นต้องมีอัลกอริทึมการสืบค้นที่สามารถสืบค้นคำศัพท์ด้วยความรวดเร็วและจัดเก็บคำศัพท์จำนวนมาก และสามารถนำพจนานุกรมดังกล่าวไปประยุกต์ใช้ในงานอื่น ๆ ได้ เช่น การตัดคำโดยใช้พจนานุกรม สำหรับการทําวินิจฉัยพจนานุกรมนี้ได้สร้างพจนานุกรมอิเล็กทรอนิกส์ภาษาไทยขึ้นมาโดยทดสอบประสิทธิภาพจากการวัดเวลาที่ใช้ในการสืบค้นคำศัพท์ และประสิทธิภาพการตัดคำโดยใช้พจนานุกรมอิเล็กทรอนิกส์ตามอัลกอริทึมที่ได้เสนอในบทที่ 5

ประสิทธิภาพการสืบค้นคำศัพท์ของพจนานุกรมอิเล็กทรอนิกส์ภาษาไทย

ประสิทธิภาพของพจนานุกรมอิเล็กทรอนิกส์ภาษาไทย พิจารณาจากเวลาที่ใช้ในการสืบค้นคำศัพท์ จากการทดสอบพบว่า เวลาที่ใช้ในการสืบค้นคำศัพท์ที่จัดเก็บในโครงสร้างข้อมูลแบบดับเบิลแฮชและโครงสร้างข้อมูลที่พัฒนาขึ้นนั้นแปรผันตามความยาวของคำศัพท์ โดยที่อัลกอริทึมการสืบค้นคำศัพท์ของโครงสร้างข้อมูลทั้ง 2 รูปแบบนั้น จะแปลงตัวอักษรแต่ละตัวของคำศัพท์เพื่อสร้างเส้นทางเดินจากโหนดรากไปยังมีลิตีโหนดหรือ เซพพาราเรทโหนดต่าง ๆ แล้วสร้างซิงเกิลสตริง เมื่อสร้างเส้นทางเพื่อเชื่อมโหนดเหล่านั้นได้สำเร็จแล้วแสดงว่ามีคำศัพท์ในพจนานุกรม แต่ถ้าไม่สามารถสร้างโหนดต่างๆ ได้หรือซิงเกิลสตริงที่ได้ไม่เหมือนกับคำศัพท์ที่ต้องการค้นแล้วแสดงว่าไม่มีคำศัพท์ในพจนานุกรม เพราะฉะนั้นการที่จะพบว่าคำศัพท์มีในพจนานุกรมหรือไม่จะต้องใช้ตัวอักษรของคำศัพท์ทุกตัว นั่นคือถึงความยาวของคำศัพท์มาก เวลาที่ใช้ในการสืบค้นคำศัพท์จะมากขึ้นด้วย เวลาที่ใช้ในการสืบค้นคำศัพท์ของอัลกอริทึมดังกล่าวเป็นไปตามฟังก์ชัน

$$O(k)$$

โดยที่ k หมายถึงความยาวของคำศัพท์ที่ต้องการสืบค้น

พจนานุกรมอิเล็กทรอนิกส์ของมหาวิทยาลัยเกษตรศาสตร์ที่นำมาเปรียบเทียบนั้น เวลาที่ใช้ในการสืบค้นคำศัพท์จากพจนานุกรมดังกล่าวคือเวลาที่ใช้ในการค้นหาจากอินเด็กซ์ เทเบิลรวมกับเวลาที่ใช้ในการสืบค้นคำศัพท์จากกลุ่มคำศัพท์โดยใช้วิธีการสืบค้นแบบทวิ (Binary Search)

แม้ว่าผลของการทดสอบจะได้ว่า ความเร็วในการสืบค้นคำศัพท์จากพจนานุกรมที่ใช้โครงสร้างข้อมูลแบบดับเบิลเอนทรีและโครงสร้างข้อมูลที่พัฒนาขึ้นด้อยกว่า พจนานุกรมของมหาวิทยาลัยเกษตรศาสตร์ แต่ความสามารถของการสืบค้นคำศัพท์ของโครงสร้างข้อมูลทั้ง 2 แบบนั้นอยู่ในเกณฑ์ที่ดี และข้อได้เปรียบของโครงสร้างข้อมูลที่พัฒนาขึ้นคือ การเพิ่มอัลกอริทึมการตัดคำโดยใช้พจนานุกรมกระทำได้โดยไม่วุ่นวาย

ประสิทธิภาพการตัดคำโดยใช้พจนานุกรมอิเล็กทรอนิกส์ภาษาไทย

การตัดคำโดยใช้พจนานุกรมอิเล็กทรอนิกส์ภาษาไทยเป็นงานวิจัยชิ้นหนึ่งที่นักวิจัยหลายท่านให้ความสนใจ วิทยานิพนธ์นี้ได้นำอัลกอริทึมการสืบค้นคำศัพท์ในพจนานุกรมอิเล็กทรอนิกส์ภาษาไทยมาพัฒนาต่อเป็นอัลกอริทึมการตัดคำโดยใช้พจนานุกรม นอกจากนี้ความสามารถการตัดคำไม่จำกัดเฉพาะผลลัพธ์เป็นคำศัพท์ที่ยาวที่สุดที่พบในพจนานุกรมเท่านั้น แต่สามารถตัดคำได้ทุกคำที่ปรากฏในพจนานุกรม ทำให้ผู้ใช้เลือกที่จะใช้คำชุดใดได้ตามความต้องการ เป็นการเพิ่มประสิทธิภาพการตัดคำทางด้านความถูกต้องของการตัดคำอีกด้วย

โครงสร้างข้อมูลสำหรับเก็บพจนานุกรมอิเล็กทรอนิกส์ภาษาไทยที่พัฒนาขึ้นโดยใช้เลือกใช้โครงสร้างข้อมูลที่เหมาะสมกับลักษณะคำไทยนั้นเป็นโครงสร้างข้อมูลประสิทธิภาพการสืบค้นทางด้านเวลาอยู่ในเกณฑ์ดี และการประยุกต์โครงสร้างข้อมูลไปใช้สำหรับการตัดคำโดยใช้พจนานุกรมอิเล็กทรอนิกส์การทำได้สะดวก และผลการตัดคำโดยใช้อัลกอริทึมนี้เป็นคำศัพท์ทุกคำที่ปรากฏในพจนานุกรมซึ่งทำให้ความถูกต้องของการตัดคำมีมากขึ้น

ข้อเสนอแนะ

วิทยานิพนธ์นี้ได้ทำการสร้างพจนานุกรมอิเล็กทรอนิกส์ภาษาไทยที่บรรจุคำศัพท์จำนวน 18,476 คำ การปฏิบัติการกับคำศัพท์นั้นจะต้องเก็บคำศัพท์ทั้งหมดไว้ในหน่วยความจำหลักทั้งหมด ถ้าจำนวนคำศัพท์มากกว่านี้แล้วจะมีผลทำให้หน่วยความจำหลักซึ่งมีจำกัดไม่สามารถรองรับข้อมูลจำนวนมากๆ ได้ แต่ถ้าสามารถปรับปรุงส่วนที่ดำเนินการเกี่ยวกับการจัดการหน่วยความจำที่ดีสามารถปฏิบัติการกับข้อมูลปริมาณมากโดยที่ไม่ทำให้ประสิทธิภาพอื่นๆ ลดลงแล้วจะเป็นการเพิ่มประสิทธิภาพการทำงานของพจนานุกรมอิเล็กทรอนิกส์ภาษาไทย

อัลกอริทึมการตัดคำศัพท์ที่พัฒนาในวิทยานิพนธ์นี้เป็นอัลกอริทึมที่ตัดคำโดยใช้พจนานุกรมเท่านั้น ถ้าต้องการให้อัลกอริทึมการตัดคำดังกล่าวสมบูรณ์ขึ้นสามารถทำโดยนำเอากฎไวยากรณ์ที่ครบถ้วนสมบูรณ์มาผสมผสานกับพจนานุกรมที่มีประสิทธิภาพที่ดีสามารถตัดคำได้เหมาะสมแล้ว จะเป็นประโยชน์อันมหาศาลสำหรับงานวิจัยเกี่ยวกับการประมวลผลด้านภาษาศาสตร์ที่จะดำเนินไปในแนวทางที่ดียิ่งขึ้น