

บทที่ 5

สรุปงานค้นคว้าแบบอิสระ

เนื่องจากภาษาไทยเป็นภาษาที่มีโครงสร้างของประโยคที่ซับซ้อนอีกทั้งยังไม่มีเครื่องหมายหรือสัญลักษณ์ที่ใช้แสดงถึงจุดสิ้นสุดของประโยคที่ชัดเจน ทำให้การสรุปใจความสำคัญสำหรับเอกสารภาษาไทยทำได้ยากและข้อสรุปที่ได้ไม่มีประสิทธิภาพเท่าที่ควร ผู้วิจัยจึงได้ทำการศึกษาวิธีการสรุปใจความสำคัญสำหรับเอกสารด้วยเทคนิคหรือวิธีการต่าง ๆ รวมไปถึงเทคนิคหรือวิธีการที่จะนำมาช่วยในการหาขอบเขตของประโยคภาษาไทยให้ถูกต้องเพื่อหาวิธีการสรุปใจความสำคัญสำหรับเอกสารภาษาไทยที่ทำให้ได้ข้อสรุปที่มีประสิทธิภาพมากที่สุด ดังนั้นงานวิจัยนี้จึงเป็นการนำเสนอวิธีการสรุปเอกสารภาษาไทยที่มีการจัดโครงสร้างโดยใช้เทคนิคการสรุปแฟร็กทัลและนำขั้นตอนวิธีวินนาวมาช่วยในการหาขอบเขตของประโยคภาษาไทยให้มีความถูกต้องมากยิ่งขึ้น โดยในการทดลองจะนำรายงานผลการดำเนินงานประจำปีพุทธศักราช 2550 ของศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ จำนวน 8 หัวเรื่อง ที่ถูกจัดทำให้อยู่ในรูปของ LaTeX มาใช้เป็นชุดข้อมูลในการทดลองเพื่อทดสอบวิธีการสรุปเอกสารภาษาไทยที่มีการจัดโครงสร้างโดยใช้เทคนิคการสรุปแฟร็กทัลตามที่ได้ออกแบบเอาไว้ จากการทดลองพบว่าข้อสรุปที่ได้มีความเที่ยงประมาณ 75%

5.1 ปัญหาและอุปสรรค

- (1) เนื้อหาบางส่วนของเอกสารต้นฉบับที่นำมาใช้ในงานวิจัยนี้มีการเว้นวรรคที่ไม่ถูกต้อง ทำให้การตัดคำและระบุชนิดของคำเกิดความผิดพลาดขึ้น เช่น คำว่า “เวลา1 สัปดาห์” ที่โปรแกรม Swath จะตัดออกมาเป็น “เวลา@NCMN1 สัปดาห์@CMTR” เป็นต้น ซึ่งจะส่งผลให้ข้อสรุปที่ได้ไม่มีประสิทธิภาพเท่าที่ควร
- (2) ประสิทธิภาพในการตัดคำของโปรแกรม Swath ยังไม่ดีเท่าที่ควร เช่น คำว่า “เนคเทค” ซึ่งโปรแกรม Swath ตัดออกมาเป็น “เนค@NPRP เท@VACT ค@NPRP” เป็นต้น ซึ่งอาจส่งผลให้ข้อสรุปที่ได้ไม่มีประสิทธิภาพเท่าที่ควร

5.2 ข้อจำกัดของการทดลอง

เอกสารต้นฉบับที่ต้องการนำมาจัดทำเป็นข้อสรุปต้องจัดทำให้อยู่ในรูปของ LaTeX เท่านั้น เนื่องจากเทคนิคการสรุปแฟร็กทัลเป็นเทคนิคที่นำโครงสร้างลำดับชั้นของเอกสารต้นฉบับมาใช้เป็นส่วนหนึ่งในการสรุปใจความสำคัญจากเอกสารต้นฉบับด้วย ดังนั้นการจัดทำเอกสาร

ต้นฉบับให้อยู่ในรูปของ LaTeX จึงมีวัตถุประสงค์เพื่อให้ผู้วิจัยสามารถตรวจสอบได้ว่าข้อความที่ปรากฏในเอกสารต้นฉบับในแต่ละส่วนเป็น Title, Section, Subsection หรือ Body ของเอกสารต้นฉบับ

5.3 แนวทางการแก้ไขปัญหา

- (1) ควรทำการตรวจสอบเนื้อหาทั้งหมดของเอกสารต้นฉบับก่อนที่จะจัดทำให้อยู่ในรูปของ LaTeX ว่ามีการเว้นวรรคที่ถูกต้องหรือไม่ เพื่อให้ข้อสรุปที่ได้มีประสิทธิภาพมากขึ้น
- (2) ควรเลือกใช้โปรแกรม เครื่องมือ หรือวิธีการอื่น ๆ ที่สามารถตัดคำจากเอกสารต้นฉบับที่อยู่ในรูปของ LaTeX และสามารถระบุชนิดของคำแต่ละคำได้เช่นเดียวกับโปรแกรม Swath แต่มีประสิทธิภาพในการตัดคำที่ดีกว่า เพื่อให้ข้อสรุปที่ได้มีประสิทธิภาพมากขึ้น

5.4 ข้อเสนอแนะ

- (1) ชุดข้อมูลฝึกหัดซึ่งจัดทำขึ้นจากคลังข้อมูลออร์คิดที่นำมาใช้ในการเรียนรู้เพื่อหาคำนำนั้ของคำแต่ละคำสำหรับงานวิจัยนี้อาจมีจำนวนที่ยังไม่เพียงพอซึ่งอาจส่งผลกระทบต่อประสิทธิภาพที่ได้ยังไม่ถูกต้องมากนัก ดังนั้นถ้ามีคลังข้อมูลใดที่สามารถนำมาสร้างเป็นชุดข้อมูลฝึกหัดที่มีจำนวนมากกว่างานวิจัยนี้ก็อาจเลือกใช้คลังข้อมูลนั้นแทนคลังข้อมูลออร์คิดได้
- (2) การใช้คนในการหาความแม่นยำของข้อสรุปที่จัดทำขึ้นจากเทคนิคการสรุปแฟร็กทัลอาจจะทำให้ผลลัพธ์ที่ได้ไม่มีความแน่นอน ดังนั้นถ้ามีโปรแกรม เครื่องมือ หรือวิธีการใดที่สามารถหาความแม่นยำเกี่ยวกับการสรุปใจความสำคัญจากเอกสารต้นฉบับได้ก็ควรนำโปรแกรม เครื่องมือ หรือวิธีการนั้นมาใช้ในการหาความแม่นยำของข้อสรุปแทนการใช้คนเพื่อให้ผลลัพธ์จากการหาความแม่นยำของข้อสรุปมีความแน่นอนมากขึ้น
- (3) นำวิธีการสรุปเอกสารภาษาไทยที่มีการจัดโครงสร้างโดยใช้เทคนิคการสรุปแฟร็กทัลที่ได้ออกแบบเอาไว้มาทำการปรับปรุงในบางขั้นตอน เช่น การนำเทคนิคหรือวิธีการอื่นมาใช้ในการหาขอบเขตของประโยคภาษาไทยแทนการใช้ขั้นตอนวิธีวินนาว การตัดคำหยุดหรือคำฟุ่มเฟือยต่าง ๆ ซึ่งปรากฏอยู่ในเอกสารต้นฉบับที่ผ่านการตัดคำแล้วออกไปก่อนที่จะทำการหาคำสำคัญของเอกสารต้นฉบับ เป็นต้น เพื่อให้เหมาะสมกับการสรุปใจความสำคัญจากเอกสารภาษาไทย ซึ่งมีโครงสร้างของประโยคที่มีความซับซ้อน