

การศึกษาปัจจัยที่มีผลต่อการย่อความข่าวภาษาไทย

The Study of Factors Affecting to Thai-News Summarization

ผศดี บุญรอด¹ และอนิราช มิ่งขวัญ²

Pudsadee Poonrawd¹ and Anirach Mingkhuang²

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษา และวิเคราะห์ถึงปัจจัยที่มีผลต่อการย่อความข่าวภาษาไทย โดยใช้วิธีวิจัยเชิงสำรวจที่เน้นวิเคราะห์โครงสร้างเนื้อหาของเอกสารต้นฉบับเพื่อใช้ในการย่อความ เครื่องมือวิจัยเป็นแบบสอบถามแบบมาตราส่วน 5 ระดับ แบ่งออกเป็น 3 ตอน ตอนที่ 1 ข้อมูลส่วนตัวของผู้ตอบแบบสอบถาม ตอนที่ 2 ขั้นตอนวิธีการย่อความข่าวและความคิดเห็นของผู้เชี่ยวชาญเกี่ยวกับปัจจัยที่มีผลต่อการย่อความข่าว และตอนที่ 3 ข้อเสนอแนะต่างๆ ในการจัดเก็บ สถิติที่ใช้ คือ ร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐานและค่าความเชื่อมั่น ผลการวิจัยพบว่า

ระดับความคิดเห็นของผู้เชี่ยวชาญที่มีต่อปัจจัยที่มีผลต่อประสิทธิภาพของระบบการย่อความข่าวภาษาไทย ประกอบด้วย 3 ด้าน คือ ด้านการนำเข้าและผลลัพธ์ที่ได้จากการย่อความ ด้านกระบวนการในการย่อความ และด้านผลลัพธ์ที่ได้จากการย่อความของระบบ โดยภาพรวมอยู่ในระดับมาก

คำสำคัญ : การวิเคราะห์เชิงเนื้อหา การย่อความข่าว เอกสารต้นฉบับ

ABSTRACT

The objective of this research was to study and analyze the factors affecting Thai-news summarization through a structure of content-based analysis of original document. The data collection instrument was a 5-rating-scale questionnaire. It was divided into three parts: personal information; simple steps of summarization and opinions of the experts on the factors affecting news summarization; and the suggestion for news summarization. The research statistics used were percentage, average score, standard deviation, and reliability statistics. The result of this study showed that there were three categories of the factors which affected Thai-news summarization: input and output, process, and outputs of Thai-news summarization system. The average opinion of the experts towards the factors affecting Thai-news summarization was high.

Keyword : Content Analysis, News Summarization, Original Document

¹ อาจารย์ ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

² พร.ต. (คอมพิวเตอร์เน็ตเวิร์ก) ผู้ช่วยศาสตราจารย์ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
รับต้นฉบับ 10 สิงหาคม 2551 รับลงพิมพ์ 10 กันยายน 2551



บทนำ

การย่อความ คือ การเก็บใจความสำคัญของเรื่องมาเรียบเรียงใหม่อย่างย่อๆ ตัดรายละเอียดที่ไม่จำเป็นออกเพื่อให้สามารถอ่านเข้าใจได้อย่างรวดเร็ว แต่ได้สาระครบถ้วนสมบูรณ์ตามเรื่องเดิม ระบบย่อความ (Text summarization system) จึงเป็นทางเลือกหนึ่งที่จะช่วยในการหาใจความสำคัญที่ปรากฏอยู่ในเอกสารต้นฉบับเพื่อย่อความเนื้อหาของเอกสารให้ผู้อ่านสามารถพิจารณาเนื้อหาของเอกสารได้จากวลีสำคัญเพียงไม่กี่ตัว และสามารถพิจารณาได้ว่าควรอ่านรายละเอียดของเอกสารต่อทั้งฉบับหรือไม่ ในปัจจุบันระบบย่อความเอกสารรองรับเอกสารภาษาอังกฤษ หรือภาษาต่างประเทศอื่นๆ ตัวอย่างเช่น เครื่องมือสรุปเอกสารอัตโนมัติของโปรแกรม Microsoft Word หรือ IBM's Intelligent text miner เป็นต้น ในส่วนของระบบการย่อความเอกสารที่รองรับภาษาไทยยังมีอยู่น้อย และอยู่ในช่วงเริ่มต้นของการพัฒนา เนื่องจากภาษาไทยมีปัญหาในการประมวลผลทางด้านภาษาในระดับคำ และระดับวลี ได้แก่ ปัญหาการตัดแบ่งคำ ปัญหาคำที่ไม่รู้จัก และไม่มีการใช้เครื่องมือทางภาษาทำให้เกิดความยุ่งยากในการประมวลผลข้อมูลด้วยเครื่องคอมพิวเตอร์

วัตถุประสงค์ของการวิจัยนี้เพื่อศึกษา และวิเคราะห์ถึงปัจจัยต่างๆ ที่มีต่อการย่อความข่าวภาษาไทย โดยเน้นวิเคราะห์โครงสร้างเนื้อหาของเอกสารต้นฉบับ ซึ่งสามารถแบ่งได้เป็น 3 ด้าน คือ ด้านส่วนนำเข้าและผลลัพธ์ที่ได้จากการย่อความ (Input and output) ด้านกระบวนการย่อความ (Process) และด้านผลลัพธ์ที่ได้จากการย่อความข่าวของระบบ (System output) เพื่อนำมาใช้ในการพัฒนาระบบย่อความข่าวภาษาไทยให้มีประสิทธิภาพ บทความนี้แบ่งออกเป็น 6 ส่วน โดยส่วนที่ 2 กล่าวถึงแนวคิดและทฤษฎีต่างๆ ที่เกี่ยวข้อง ส่วนที่ 3 กล่าวถึงวิธีดำเนินการวิจัย ส่วนที่ 4 กล่าวถึงผลที่ได้จากการวิจัย ส่วนที่ 5 สรุปผลการวิจัย และส่วนที่ 6 เป็นข้อเสนอแนะของงานวิจัย

แนวคิดและทฤษฎีที่เกี่ยวข้อง

1. การย่อความเอกสาร (Text summarization)

การย่อความเอกสารเป็นการสร้างตัวแทนของเอกสารต้นฉบับให้อยู่ในรูปแบบของคำ (Word) หรือวลีที่เรียงต่อกันเพื่อแสดงใจความสำคัญของเนื้อหาในเอกสารต้นฉบับทั้งนี้ข้อมูลบางส่วนอาจถูกตัดออก โดยผ่านกระบวนการคัดสรรและย่อความเอกสาร ซึ่งการย่อความเอกสารจะช่วยให้ผู้อ่านสามารถมองเห็นภาพรวมของเอกสารตัวจริงได้พอสมควร ทำให้อ่านได้ง่าย รวดเร็ว และสามารถพิจารณาได้ว่าควรอ่านเอกสารต่อทั้งฉบับหรือไม่ ระบบการย่อความ (Text summarizer) เป็นระบบที่ให้ผู้เลือกใช้เอกสารนำเข้าหรือเอกสารต้นฉบับที่ต้องการย่อความ จากนั้นระบบจะสร้างผลลัพธ์ของการย่อความให้โดยอัตโนมัติตามความต้องการของผู้ใช้งาน ซึ่งในกระบวนการย่อความจะต้องอาศัยหลักการประมวลผลของภาษารวมชาติที่ต้องมีการวิเคราะห์ทางความหมาย (Semantic) มีการตีความทางภาษา แบ่งกลุ่มเนื้อหา และสร้างรูปแบบของภาษาขึ้นมา โดยมีการนำแนวคิดต่างๆ เข้ามาช่วยในการวิเคราะห์คำ วลี และรูปประโยค เช่น เทคนิคการค้นคืนข้อมูล (Information retrieval) เครื่องจักรกลเรียนรู้ (Machine learning) หรือหลักการทางสถิติบนพื้นฐานของการนับจำนวนคำ และการแบ่งกลุ่มข้อมูล

1. เทคนิคการย่อความ

กระบวนการย่อความเอกสารมีหลายแนวทาง ตัวอย่างเช่น Hovy และ Lin ได้นำเสนอกระบวนการย่อความไว้ 2 แนวทาง คือ กระบวนการค้นคืนสารสนเทศ (Information retrieval) และกระบวนการสกัดสารสนเทศ (Information extraction) หรือ Carbonell และ Goldstein ได้นำเสนอเทคนิคของการย่อความจากคำเดียว กลุ่มคำ หรือย่อความในลักษณะเชิงแนวคิด (Concept) โดยการย่อความด้วยข้อความที่สั้นและดีที่สุด (Best passage) หรือนำหลายๆ ประโยคมาเรียงต่อกัน หรือย่อความด้วยการสร้างประโยคใหม่ เป็นต้น โดยเทคนิคการย่อความภาษาไทยอาจเลือกปฏิบัติได้ 3 แนวทาง ตามความเหมาะสมของเนื้อหาที่สรุปและประโยคที่จะนำไปใช้ ซึ่งมีรายละเอียดดังนี้

2. เทคนิคการวิเคราะห์ตำแหน่งความคิดสำคัญ

เทคนิคนี้อาศัยความรู้ในเรื่องโครงสร้างของงานเขียน โดยย่อหน้าหนึ่งประกอบด้วยประโยคใจความสำคัญ และ ประโยคสนับสนุน ซึ่งส่วนใหญ่งานเขียนที่เป็นระบบจะวางความคิดสำคัญไว้ที่ต้นย่อหน้า หรือท้ายย่อหน้า หรือบางครั้ง อาจวางความคิดสำคัญไว้ทั้งต้นและท้ายย่อหน้า หรือวางไว้ กลางย่อหน้า ทั้งนี้การวางความคิดสำคัญไว้ทั้งต้นและท้ายย่อ หน้าในตัวอย่างงานเขียนของไทยจะพบเห็นได้มาก แต่การ วางความคิดสำคัญไว้กลางย่อหน้าจะมีเป็นตัวอย่างให้พบเห็น ได้น้อย

3. เทคนิคการถาม - ตอบ

เทคนิคนี้อาจใช้ในกระบวนการอ่านเพื่อคิดสรุป โดย ทำการตั้งคำถามไว้หลายๆ ประเภท เช่น คำถามให้สังเกต คำถามให้อธิบาย คำถามสรุป เป็นต้น ซึ่งคำถามที่เหมาะสมแก่ การสรุปความ คือ สามารถตอบได้ว่ามีใครทำอะไร ที่ไหน และเมื่อใด

4. เทคนิคการหาคำหลัก

เทคนิคนี้จะต้องหาคำสำคัญที่อยู่ในประโยคใจความ สำคัญ ประโยคสนับสนุนหลัก และประโยคสนับสนุนรอง ซึ่ง วิธีหาคำหลักที่ได้จากประโยคใจความสำคัญและประโยค สนับสนุน อาจทำการขีดเส้นใต้เฉพาะคำสำคัญในประโยค ใจความสำคัญ ประโยคสนับสนุนหลัก และประโยคสนับสนุน รอง โดยจะต้องคำนึงถึงหลักไวยากรณ์ จากนั้นทำการบันทึก ย่อ

รูปแบบผลลัพธ์ที่ได้จากการย่อความประกอบด้วย เนื้อหา 2 ส่วน คือ แหล่งที่มาของเรื่อง เพื่อบอกรายละเอียด รูปแบบของแหล่งที่มา ซึ่งจะแตกต่างกันตามลักษณะของ เอกสาร และสาระของเรื่องที่ย่อความ เพื่อแสดงสาระของ เรื่องที่ย่อความที่ได้มาจากการอ่าน และเก็บใจความสำคัญของเรื่อง ที่อ่าน ขนาดความยาวของเรื่องไม่มีกฎเกณฑ์ตายตัว ขึ้นอยู่กับสาระสำคัญของเรื่อง โดยตัวอย่างการเขียนย่อความ ขาวสามารถแสดงได้ดังตัวอย่างรูปแบบการย่อความขาว

ย่อความขาวเรื่อง สถานการณ์โรคเบาหวานใน ประเทศไทย จาก ไทยรัฐ วันที่ 11 พฤศจิกายน 2548

ความว่า

องค์การอนามัยโลกระบุว่าโรคเบาหวานเป็นโรคที่ อันตรายสูงยิ่งกว่าโรคเอดส์ ส่วนสถานการณ์โรคเบาหวาน ในประเทศไทยพบว่าโรคนี้นักเข้ามาในกลุ่มของเด็กมากขึ้น ด้านน.พ.สุริยเดว ทรีปาตี โฆษกเครือข่ายเด็กไทยไม่กินหวานกล่าวว่ามีคนไข้เด็กป่วยด้วยโรคเบาหวานเพิ่มมากขึ้น ทุกปี และมีอายุน้อยลง

เรื่อย โดยสัญญาณบ่งบอกถึงโรคเบาหวานในเด็ก คือ อ้วนจนผิวหนังรอบคอ และรักแร้ดำเป็นมัน มีอาการปัสสาวะ บ่อย หิวบ่อย กินจุ ซาตามปลายมือปลายเท้า หรือมีพ้อ แม่ ปู่ ย่า ตา ยายเป็นเบาหวาน

ตัวอย่าง รูปแบบการย่อความขาว

3. งานวิจัยที่เกี่ยวข้องกับการย่อความเอกสาร

งานวิจัยที่เกี่ยวข้องกับการย่อความเอกสารภาษา อังกฤษในปัจจุบัน มีการศึกษาและพัฒนาอย่างต่อเนื่อง แต่สำหรับภาษาไทยยังมีน้อย หรืออยู่ในช่วงเริ่มต้นการ พัฒนา ซึ่งผู้วิจัยได้ศึกษาและสรุปถึงงานวิจัยต่างๆ ที่ เกี่ยวข้องดังต่อไปนี้

งานวิจัยที่เกี่ยวข้องกับการย่อความภาษาอังกฤษมี การศึกษา และพัฒนาอย่างต่อเนื่อง ตัวอย่างเช่น การย่อ ความเอกสารโดยใช้ Word-cluster และ Machine learning Ranking algorithm [8] ซึ่งระบบสามารถย่อความ ได้อย่างมีประสิทธิภาพ แต่ยังไม่รองรับการย่อความแบบ Multi-document จากผลการประเมินหาประสิทธิภาพพบว่า ระบบทำงานได้ดีใน Mean coverage, Mean length adjusted coverage และ Quality question score เมื่อ เปรียบ เทียบกับระบบอื่นๆ งานวิจัยของ Turney ได้พัฒนาระบบ Gen Ex ที่มีแนวคิดในการสกัดวลีสำคัญจากเอกสาร โดยใช้ 2 เทคนิคร่วมกัน คือ Genitor และ Extractor จาก ผลการทดลองพบว่า Gen Ex สามารถเลือกวลีสำคัญได้ตรงกับ ผู้เขียนเอกสารประมาณ 20 เปอร์เซ็นต์และระบบ Kea เป็นงานวิจัยที่มีแนวคิดในการสกัดวลีสำคัญจากเอกสารโดย



ใช้หลักการของเบย์ (Naïve bayes algorithm) ในการเรียนรู้ เพื่อสร้างแบบจำลองสำหรับแยกกลุ่มวลีในเอกสาร โดยพิจารณาจาก 2 ลักษณะสำคัญ คือ คำนำหน้าของวลี และ ตำแหน่งที่พบวลีครั้งแรกในเอกสาร

ซึ่งใช้หลักการของเบย์เรียนรู้ จากผลการทดลองความถูกต้องของวลีสำคัญที่เลือกโดยใช้เทคนิค Kea ใกล้เคียงกับวลีสำคัญที่เลือกโดย Gen Ex แต่การคำนวณมีความซับซ้อนน้อยกว่า Gen Ex

ตัวอย่างงานวิจัยการย่อความภาษาจีน ได้แก่ งานวิจัยของ Lee ได้นำเสนอการย่อข่าว ประจำวัน โดยใช้หลักการของฟัซซี-ออนโทโลยี (Fuzzy-ontology) แต่ระบบมีข้อจำกัดในเรื่องของคำบางคำที่มีความหมายไม่ชัดเจน ส่วนงานวิจัยของ Chia-Wei และคณะ ได้พัฒนาระบบย่อความภาษาจีนสำหรับข่าวหรือบทความทางธุรกิจ โดยใช้ฐานข้อมูลออนโทโลยีสำหรับวิเคราะห์หัวข้อหลักของบทความหลังจากระบุหัวข้อหลักและความสัมพันธ์แล้ว จะจัดลำดับย่อหน้าโดยใช้ความเกี่ยวข้องกันของหัวข้อหลักและย่อหน้า ซึ่งขึ้นอยู่กับความเรียงลำดับ และใช้อัตราส่วนของย่อหน้าเพื่อสร้างสรุปข้อความ โดยผลการทดลองเมื่อเปรียบเทียบกับวิธี Term-frequency พบว่าการนำออนโทโลยีมาใช้จะช่วยให้การหาหัวข้อย่อได้ดีขึ้น และได้ผลในด้านความถูกต้องใกล้เคียงกับวิธี Term-frequency ในการเลือกย่อหน้า

ตัวอย่างงานวิจัยที่เกี่ยวข้องกับการย่อความภาษาไทย ได้แก่ การย่อความเอกสารภาษาไทยโดยการวิเคราะห์หาความหมายที่แอบแฝง (Latent semantic) ซึ่งมีข้อจำกัดในการหาขอบเขตคำ วลี และการตัดคำ ผู้วิจัยได้ให้ข้อเสนอแนะไว้ว่าควรนำโมเดลทางภาษามาใช้ร่วมกับการประมวลผลของภาษาธรรมชาติมาช่วยในการประมวลผล เพื่อลดข้อผิดพลาดในการตัดคำ และเพิ่มประสิทธิภาพในการย่อความมากยิ่งขึ้น งานวิจัยของซูลีรัตน์และคณะ ที่ได้แนะนำเสนอ Practical approach สำหรับการสกัดข้อความจากเอกสารต้นฉบับ เพื่อย่อความโดยเริ่ม ต้นแบ่งกลุ่มคำสำคัญ ค้นหาความสัมพันธ์ของ Text span และใช้วิธี Local properties สำหรับแบ่งกลุ่มคำสำคัญในแต่ละย่อหน้า และวิธี Global สำหรับพิจารณาความสัมพันธ์ของทุกๆ ย่อหน้าจากผลการ

ทดลองพบว่าการย่อความที่มีความยาว 20 % จากเอกสารต้นฉบับ มีประสิทธิภาพสูงกว่าเมื่อนำไปเทียบกับ Auto summarize ในโปรแกรม Microsoft proofing tools for office แต่งานวิจัยนี้มีข้อจำกัดในการย่อความเอกสาร หากเอกสารมีความยาวน้อยเกินไป อาจทำให้ผลลัพธ์ในการย่อความผิดพลาด และผลลัพธ์ในการย่อความยังพบคำที่ไม่จำเป็นรวมอยู่ด้วย

วิธีดำเนินการวิจัย

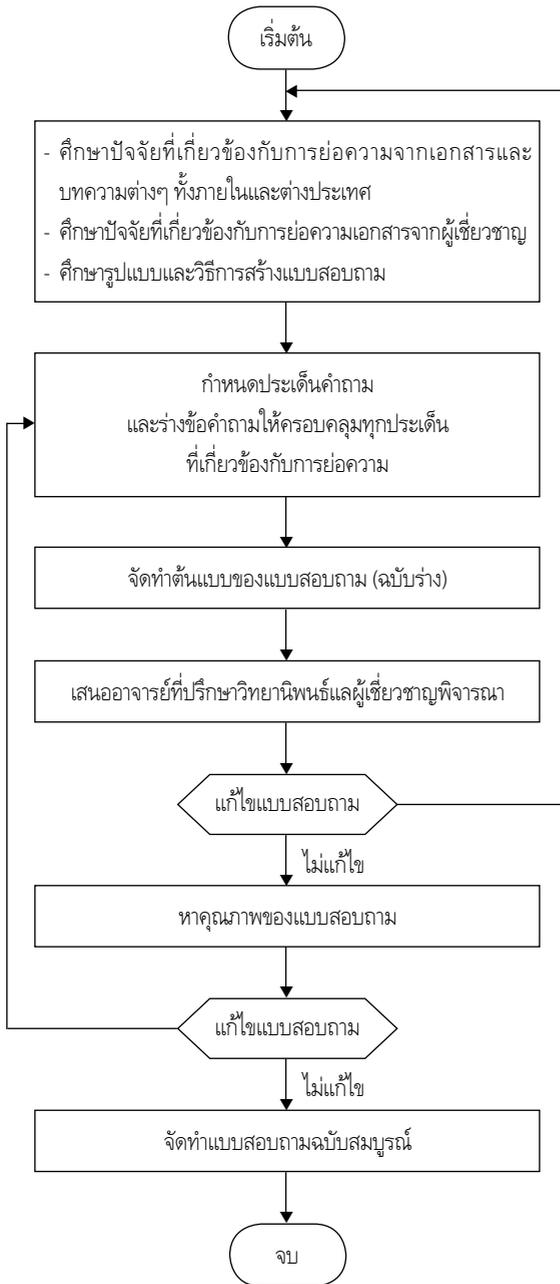
การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษา และวิเคราะห์ถึงปัจจัยที่มีผลต่อประสิทธิภาพของการย่อความข่าวภาษาไทย โดยผู้วิจัยใช้วิธีวิจัยเชิงสำรวจ (Survey research) โดยกำหนดขั้นตอนในการวิจัยตามลำดับดังนี้

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการวิจัยสามารถแบ่งได้เป็น 2 กลุ่มได้แก่ กลุ่มแรกสำหรับหาความเชื่อมั่นของแบบ สอบถาม ซึ่งเป็นครูผู้สอนวิชาภาษาไทย จำนวน 30 คน และกลุ่มที่สองเป็นผู้เชี่ยวชาญในการประเมินแบบ สอบถาม ซึ่งเป็นครูภาษาไทยที่มีประสบการณ์ในการย่อความ จำนวน 66 คน

เครื่องมือการวิจัย

ผู้วิจัยใช้แบบสอบถามเป็นเครื่องมือในการวิจัย โดยแบบสอบถามเป็นแบบมาตราส่วนประมาณค่า 5 ระดับ ที่แบ่งออกเป็น 3 ตอน คือ ตอนที่ 1 ข้อมูลส่วนตัวของผู้ตอบแบบสอบถาม ตอนที่ 2 ขั้นตอนวิธีการย่อความข่าวและความคิดเห็นของผู้เชี่ยวชาญเกี่ยวกับปัจจัยที่มีผลต่อการย่อความข่าว และตอนที่ 3 ข้อเสนอแนะต่างๆ ในแต่ละด้าน ในการสร้างแบบสอบถามฉบับสมบูรณ์ ผู้วิจัยได้กระทำอย่างเป็นลำดับขั้นตอนดังแผนภูมิที่ 1



แผนภูมิที่ 1 ขั้นตอนการสร้างแบบสอบถาม

ในการทำแบบสอบถามผู้วิจัยได้ศึกษาข้อมูลต่างๆ ที่จำเป็นในการทำแบบสอบถาม เพื่อกำหนดประเด็นคำถาม และร่างข้อคำถามให้ครอบคลุมจนได้แบบสอบถามฉบับร่าง จากนั้นนำไปเสนอให้อาจารย์ที่ปรึกษาและผู้เชี่ยวชาญทางด้านภาษาไทยพิจารณา เพื่อทำการปรับแก้ไขแบบสอบถาม

และนำแบบสอบถามดังกล่าวไปให้กลุ่มตัวอย่างที่เป็นครูภาษาไทยจำนวน 30 คน ทดสอบเพื่อหาความเชื่อมั่นของแบบสอบถาม ซึ่งในส่วนนำเข้าข้อมูลและผลลัพธ์ที่ได้จากการย่อความได้ค่า Cronbach's alpha มีค่าเท่ากับ 0.80 ด้านกระบวนการย่อความได้ค่า Cronbach's alpha มีค่าเท่ากับ 0.70 ด้านผลลัพธ์ที่ได้จากการย่อความข่าวของระบบได้ค่า Cronbach's alpha มีค่าเท่ากับ 0.79 และได้ค่า Cronbach's alpha โดยรวมซึ่งหาจากปัจจัยทุกด้านของแบบสอบถามมีค่าเท่ากับ 0.81 แสดงว่าแบบสอบถามที่ใช้เป็นเครื่องมือวัดมีความน่าเชื่อถือสูง

การเก็บรวบรวมและวิเคราะห์ผลข้อมูล

เมื่อได้แบบสอบถามฉบับสมบูรณ์แล้ว ผู้วิจัยได้นำแบบสอบถามไปให้ผู้เชี่ยวชาญ ซึ่งเป็นกลุ่มตัวอย่างที่ 2 ที่เป็นครูภาษาไทยที่มีประสบการณ์ในการย่อความ จำนวน 66 คน ทำการประเมินแบบสอบถาม เมื่อเก็บรวบรวมข้อมูลทั้งหมดแล้ว ได้นำมาตรวจสอบความถูกต้องของข้อมูล ทำการกำหนดชุดตัวแปรและรหัสข้อมูล เพื่อป้อนข้อมูลลงในโปรแกรมคอมพิวเตอร์สำเร็จรูปจากนั้นทำการวิเคราะห์ข้อมูลโดยใช้สถิติ คือ ร้อยละ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน

ผลการวิจัย

ผลการวิจัยมีดังนี้

จากการวิเคราะห์ข้อมูลด้วยโปรแกรมคอมพิวเตอร์แสดงให้เห็นถึงข้อมูลส่วนตัวของผู้เชี่ยวชาญ ซึ่งส่วนใหญ่จบการศึกษาระดับปริญญาตรี ร้อยละ 92.42 และจบการศึกษาระดับปริญญาโท ร้อยละ 6.06 ส่วนใหญ่มีประสบการณ์ทำงานเป็นครูสอนภาษาไทยมากกว่า 15 ปีขึ้นไป ร้อยละ 92.42 และมีประสบการณ์ทำงานต่ำกว่า 5 ปี ร้อยละ 6.06 ส่วนใหญ่มีประสบการณ์ในการย่อความมากกว่า 10 ปีขึ้นไป ร้อยละ 80.30 ประสบการณ์ในการย่อความ 7 ปีแต่ไม่ถึง 10 ปี ร้อยละ 7.58 ประสบการณ์ในการย่อความต่ำกว่า 1 ปี ร้อยละ 4.55 ประสบการณ์ในการย่อความ 1 ปีแต่ไม่ถึง 3 ปี ร้อยละ 3.03 และประสบการณ์ในการย่อความ 4 ปีแต่ไม่ถึง 6 ปี ร้อยละ 1.52



ในการพัฒนาระบบการย่อความเอกสารภาษาไทยนั้นทำได้ยาก เนื่องจากภาษาไทยมีปัญหาในการประมวลผลในระดับคำ และระดับวลีผู้วิจัยจึงต้องทำการศึกษา และวิเคราะห์ถึงปัจจัยต่างๆ ที่มีผลต่อประสิทธิภาพของการย่อความเอกสารภาษาไทย จากผลการวิจัยที่ใช้กรณีศึกษาเป็นข่าววิทยากรและข่าวเศรษฐกิจ และทำการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่าง คือ ครูภาษาไทยที่มีประสบการณ์ในการย่อความ จำนวน 66 คน โดยเน้นวิเคราะห์จากโครงสร้างของเนื้อหาเอกสารต้นฉบับ พบว่า ปัจจัยที่ส่งผลต่อการย่อความข่าวภาษาไทย แบ่งได้เป็น 3 ด้าน คือ ด้านส่วนนำเข้าและผลลัพธ์ที่ได้จากการย่อความ ด้านกระบวนการในการย่อความ และด้านผลลัพธ์ที่ได้จากการย่อความของระบบ ซึ่งมีระดับความคิดเห็นโดยภาพรวมอยู่ในระดับมาก โดยในแต่ละด้านมีปัจจัยต่างๆ ดังนี้ ด้านส่วนนำเข้าข้อมูลและผลลัพธ์ที่ได้จากการย่อความ (Input and output) มีปัจจัยย่อย ได้แก่ โครงสร้างของเอกสารต้นฉบับ ขนาดของเอกสารต้นฉบับ ลักษณะความต้องการใช้งานเอกสาร หัวข้อเรื่องที่ย่อความ คำหยุด รูปแบบประโยคที่ปรากฏอยู่ในผลลัพธ์ และขนาดของผลลัพธ์ที่ได้จากการย่อความ ซึ่งผู้ใช้ควรที่จะสามารถกำหนดรูปแบบของผลลัพธ์ (Format) และขนาดความยาวของเอกสารที่ต้องการได้ เนื่องจากขนาดของเอกสารจะมีผลต่อการย่อความให้ได้ความหมาย ในด้านกระบวนการย่อความ (Process) มีปัจจัยย่อย ได้แก่ โครงสร้าง รูปแบบของภาษา สถานการณ์ เป้าหมาย ลักษณะการนำผลลัพธ์ไปใช้งาน ที่อาจย่อความจากในเอกสารต้นฉบับเดิม หรือสร้างเป็นเอกสารใหม่ และผู้เกี่ยวข้องที่ปรากฏอยู่ในเอกสาร (Situation) ซึ่งอาจบ่งบอก หรือไม่ได้บ่งบอกผู้กระทำสถานที่ และเวลาที่เหตุการณ์ปรากฏ ในด้านผลลัพธ์ที่ได้จากการย่อความของระบบ (System output) มีปัจจัยย่อย ได้แก่ ความสามารถกำหนดรายละเอียดผลลัพธ์ของผู้ใช้งาน ความสามารถในการกำหนดน้ำหนักของคำ ขั้นตอนวิธีการในการย่อความ และรูปแบบของผลลัพธ์ที่ได้จากการย่อความ ซึ่งปัจจัยต่างๆ เหล่านี้ ผู้พัฒนาระบบจะต้องให้ความสำคัญและคำนึงถึง เพื่อสามารถพัฒนาระบบการย่อความได้อย่างมีประสิทธิภาพ ตรงกับความต้องการของผู้ใช้งาน และสามารถ

รองรับการย่อความที่มีขอบเขตของเนื้อหาต่างกัน ได้อย่างถูกต้องสมบูรณ์

ข้อเสนอแนะเพื่อนำผลการวิจัยไปใช้

1. ในการย่อความเอกสารควรวิเคราะห์แยกแยะได้ว่าในเอกสารมีใครทำอะไร ที่ไหน อย่างไร
2. การพัฒนาระบบย่อความเอกสารอาจจะเหมาะสมกับงานบางประเภท เช่น งานที่มีข้อมูลจำนวนมาก งานวิชาการ หรืองานประชาสัมพันธ์ ซึ่งอาจขึ้นอยู่กับวัตถุประสงค์ของงาน หรือเหมาะกับงานบางประเภทเท่านั้น

เอกสารอ้างอิง

- แก้วตา กรุงวรงค์. (2544). **ภาษาไทยธุรกิจ**. กรุงเทพมหานคร: บริษัทเพิสท์ ออฟเซท (1993) จำกัด.
- สุลธิรัตน์ จรัสกุลชัย และชนวนาคัย กริ่งไกร. (2545). **การพิสูจน์หัวหัวเรื่องที่มีนัยสำคัญ และการย่อความของข้อความภาษาไทยแบบอัตโนมัติ** ภาควิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยเกษตรศาสตร์.
- ชนุตศักดิ์ ธัญญสิริ. (2543). **การออกแบบและพัฒนาระบบต้นแบบสำหรับการสืบค้นข้อมูลสารสนเทศภาษาไทยด้วยดัชนีหลายระดับ**. วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์.
- นฤมล ปราชญ์โยธิน. (2536). **ทวิศักดิ์ กอนันตกุล และเปรมิน จินดาวิมลเลิศ. ธิซอร์สกับระบบสารสนเทศ**. กรุงเทพฯ : ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ.
- อัษฎางค์ แต่งไทย. (2548). **การย่อความเอกสารภาษาไทยโดยการวิเคราะห์หาความหมายที่แอบแฝง**. วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์.
- Carbonell, J. and J. Goldstein. (1998). **The use of MMR, diversity based reranking for reordering documents and producing summaries**. In Proceedings of SIGIR.

- Chia - Wei, W. and L. Chao-Lin. (2003). **Ontology-based Text Summarization for Business News Articles**. National Chengchi University Taiwan.
- Frank, E. *et. al.* (1999). Domain - Specific Key phrase Extraction, **In Proceedings of the 16th International Joint Conference on Artificial Intelligence**. Morgan Kaufmann Publishers, San Francisco, CA.
- Hovy E. and C. Lin. (1997). **Automated Text Summarization in SUMMARIST**. pp. 18-24. In Proceedings of the ACL Workshop on Intelligence Text Summarization.
- Jaruskulchai, C. and C. Kruengkrai. (2003). **A Practical Text Summarizer by Paragraph Extraction for Thai**. The Sixth International Workshop on Information Retrieval with Asian Languages (IRAL - 2003), July 7.
- Lee, C. S., Z. W. Jian and L. K. Huang. (2005). Fuzzy ontology and its application to news Summarization. **In IEEE Transactions on systems. Cybernetics**, Vol.35, No.5: 859-880.
- Massih R.A.*et.al.*,(2005). **Automatic Text Summarization Based on Word-Clusters and Ranking Algorithms**. ECIR 2005, LNCS 3408, pp. 142-156.
- Turney, P. D. (1999). **Learning to Extract Key phrases from Text**. Technical Report ERB - 1057, National Research Council, Institute for Information Technology.

