



# การวิเคราะห์ปัจจัยเชิงยืนยันที่มีผลต่อการย่อความข่าวภาษาไทย

ผุสดี บุญรอด\* อนิราช มิ่งขวัญ\*\* และศจีมาศ ณ วิเชียร\*\*\*

## บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษา วิเคราะห์ปัจจัยที่มีผลต่อการย่อความข่าวภาษาไทย และวิเคราะห์องค์ประกอบเพื่อหาความสัมพันธ์เชิงโครงสร้าง ของปัจจัยที่มีผลต่อการย่อความข่าวภาษาไทย โดยเน้นการวิเคราะห์โครงสร้างของเอกสารต้นฉบับ เครื่องมือที่ใช้ในการวิจัยเป็นแบบสอบถามแบบมาตราส่วน 5 ระดับ ที่แบ่งเป็น 3 ตอน คือ ตอนที่ 1 ข้อมูลส่วนตัวของผู้ตอบแบบสอบถาม ตอนที่ 2 ขั้นตอนวิธีการย่อความข่าว และความคิดเห็นของผู้ตอบแบบสอบถามเกี่ยวกับปัจจัยที่มีผลต่อการย่อความข่าว และตอนที่ 3 ข้อเสนอแนะต่างๆ ทำการจัดเก็บ และวิเคราะห์ข้อมูลโดยใช้โปรแกรม SPSS for Windows และวิเคราะห์ปัจจัยเชิงยืนยันที่มีผลต่อการย่อความข่าวภาษาไทย โดยใช้โปรแกรม LISREL ผลการวิจัยสรุปได้ว่า ปัจจัยที่มีผลต่อประสิทธิภาพของการย่อความข่าวภาษาไทยประกอบด้วย 3 ด้าน คือ ด้านส่วนนำเข้าของเอกสารต้นฉบับ ด้านกระบวนการย่อความ และผลลัพธ์ที่ได้จากการย่อความเอกสาร ซึ่งผู้เชี่ยวชาญมีระดับความคิดเห็น โดยภาพรวมอยู่ในระดับมาก ( $\bar{X} = 4.03$ , S.D. = 0.99) และผลการวิเคราะห์ปัจจัยเชิงยืนยัน ที่มีผลต่อการย่อความข่าวพบว่า แต่ละด้านมีความสอดคล้องกับข้อมูลเชิงประจักษ์ดี ซึ่งแตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

**คำสำคัญ:** การวิเคราะห์องค์ประกอบ การวิเคราะห์ปัจจัยเชิงยืนยัน การย่อความข่าว เอกสารต้นฉบับ

## 1. บทนำ

ในปัจจุบันเป็นยุคข้อมูลข่าวสาร ที่ข้อมูลและบริการจำนวนมากถูกนำเสนอผ่านสื่ออินเทอร์เน็ต แต่ปัญหาที่ผู้ใช้กำลังประสบอยู่ในขณะนี้ คือ ข้อมูลและบริการที่มีอยู่มากมายนั้น กำลังเกิดภาวะท่วมล้น มีจำนวนมากเกินไป [1] เพื่อช่วยลด

ระยะเวลาในการอ่าน และการทำความเข้าใจเอกสารจำนวนมาก ระบบย่อความเอกสาร (Text Summarization System) จึงเป็นทางเลือกหนึ่ง ที่ช่วยในการหาข้อสำคัญที่ปรากฏอยู่ในเอกสารต้นฉบับ (Original Documents) เพื่อย่อความเนื้อหาของเอกสาร ให้ผู้อ่านสามารถพิจารณาเนื้อหาของเอกสารได้จากข้อสำคัญเพียงไม่กี่ตัว และสามารถพิจารณาได้ว่าควรอ่านรายละเอียดของเอกสารต่อทั้งฉบับหรือไม่ ระบบการย่อความเอกสารที่รองรับภาษาไทยในปัจจุบันยังมีอยู่น้อย และอยู่ในช่วงเริ่มต้นของการพัฒนา [2] เนื่องจากภาษาไทยมีปัญหาการประมวลผลทางด้านภาษาในระดับคำ และระดับวลี [3] ได้แก่ ปัญหาการตัดแบ่งคำ ปัญหาคำที่ไม่รู้จัก และไม่มีการใช้เครื่องมือทางภาษา ทำให้เกิดความยุ่งยากในการประมวลผลข้อมูลด้วยเครื่องคอมพิวเตอร์

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษา วิเคราะห์ปัจจัยที่มีผลต่อการย่อความข่าว และวิเคราะห์องค์ประกอบเพื่อหาความสัมพันธ์เชิงโครงสร้าง ของปัจจัยที่มีผลต่อการย่อความข่าวภาษาไทย โดยเน้นวิเคราะห์โครงสร้างเนื้อหาของเอกสารต้นฉบับ [4] ที่สามารถแบ่งได้เป็น 3 ด้าน คือ ด้านส่วนนำเข้าของเอกสารต้นฉบับ (Input) ด้านกระบวนการย่อความ (Process) และด้านผลลัพธ์ที่ได้จากการย่อความเอกสาร (Output) เพื่อนำมาใช้ในการพัฒนาระบบย่อความข่าวภาษาไทยให้มีประสิทธิภาพ โดยบทความนี้แบ่งออกเป็น 5 ส่วน ส่วนที่ 2 กล่าวถึงแนวคิดและทฤษฎีต่างๆ ที่เกี่ยวข้อง ส่วนที่ 3 กล่าวถึงวิธีดำเนินการวิจัย ส่วนที่ 4 กล่าวถึงผลที่ได้จากการวิจัย และส่วนที่ 5 สรุปผลและข้อเสนอแนะของงานวิจัย

## 2. แนวคิดและทฤษฎีที่เกี่ยวข้อง

### 2.1 การย่อความเอกสาร

การย่อความเอกสารเป็นการเก็บใจความสำคัญ ในเรื่องที่ต้องการย่อมาเรียบเรียงใหม่อย่างย่อ ๆ โดยตัดข้อความที่ไม่

\* คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ \*\* วิทยาลัยเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

\*\* คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

สำคัญ และไม่จำเป็นออกไป เขียนขึ้นใหม่แต่ความหมายหรือสาระสำคัญของเรื่องยังมีอยู่คงเดิมไม่ขาดหาย [5] เพื่อให้สามารถอ่านเข้าใจได้อย่างรวดเร็ว แต่ได้สาระครบถ้วนสมบูรณ์ตามเรื่องเดิม โดยระบบย่อความเอกสาร (Text Summarization System) เป็นระบบที่ช่วยในการหาใจความสำคัญที่ปรากฏอยู่ในเอกสารต้นฉบับ เพื่อสรุปเนื้อหาของเอกสาร ให้ผู้อ่านสามารถพิจารณาเนื้อหาของเอกสารได้จากวลีสำคัญเพียงไม่กี่ตัวของย่อความเอกสารที่พบในชีวิตประจำวัน เช่น หัวข้อข่าว (Headlines) ในกรอบหนังสือพิมพ์ หรือการแสดงผลลัพธ์ของโปรแกรมค้นหาข้อมูล (Search Engine) ที่แสดงผลลัพธ์เป็นการสรุปใจความสำคัญของคำขอ (Query) ที่ช่วยให้ผู้ใช้ตัดสินใจเลือกเว็บไซต์ที่ตรงตามความต้องการ เป็นต้น

แต่ในการพัฒนาระบบย่อความนั้น เป็นปัญหาที่ยากในการพัฒนา [6] เนื่องจากต้องอาศัยการประมวลผลของภาษาธรรมชาติ (Natural Language Processing: NLP) ที่ต้องมีการวิเคราะห์ตีความทางภาษา แบ่งกลุ่มเนื้อหา และสร้างรูปแบบของภาษาขึ้นมา จึงไม่สามารถพัฒนาเป็นระบบย่อความอัตโนมัติได้อย่างสมบูรณ์การย่อความเอกสารมีวัตถุประสงค์เพื่อลดขนาดของเอกสาร โดยแสดงเพียงใจความสำคัญของเอกสาร เพื่อสามารถอ่านเอกสารได้เพิ่มมากขึ้นในเวลาเท่าเดิม ซึ่งกระบวนการสำหรับการย่อความเอกสารสามารถแบ่งออกได้เป็นหลายแนวทาง เช่น Hovy and Lin [7] ได้นำเสนอกระบวนการย่อความไว้ 2 แนวทาง คือ กระบวนการค้นคืนสารสนเทศ (Information Retrieval) และกระบวนการสกัดสารสนเทศ (Information Extraction) หรือ Carbonell and Goldstein [8] ได้นำเสนอเทคนิคของการย่อความด้วยคำ ๆ คำเดียว จากคำหลาย ๆ คำหรือการย่อความโดยใช้ Word-Cluster และ Machine Learning Ranking Algorithm [9] ที่ใช้การวิเคราะห์คำและการจัดลำดับ โดยนำวิธีการเครื่องจักรเรียนรู้มาประยุกต์ใช้งาน หรือการย่อความด้วยคำในลักษณะเชิงแนวคิด (Concept) การย่อความด้วยข้อความที่สั้นและดีที่สุด (Best Passage) การย่อความจากหลายประโยคที่เรียงต่อกัน หรือการย่อความด้วยการสร้างประโยคใหม่เป็นต้น หรือการย่อความแบบสาระสังเขป (Abstract Summarization) ที่ต้องอาศัยความรู้ทางด้านการประมวลผลของภาษาธรรมชาติที่มีการย่อความโดยสร้างผลลัพธ์เป็นเนื้อหาสรุปขึ้นมาใหม่ ให้ได้ใจความสำคัญ และสามารถมองเห็นภาพรวมของเอกสารต้นฉบับได้ทั้งหมด โดยในส่วนของโครงสร้างภาษาไทยนั้น มีรูปแบบการเขียนบรรยายเป็นตัวอักษรต่อเนื่องกัน ไม่มี

สัญลักษณ์บ่งชี้จุดสิ้นสุดของคำ ซึ่งรูปแบบการแบ่งประโยคจะไม่ปรากฏเด่นชัดเหมือนในภาษาอังกฤษ ที่ใช้จุดในการจบประโยค (Full Stop) ซึ่งทำให้การประมวลผลข้อความภาษาไทยมีความยาก และซับซ้อน

ตัวอย่างงานวิจัยที่เกี่ยวข้องกับการประมวลผลภาษาไทย และการย่อความภาษาไทย ได้แก่ งานวิจัยของซูลีร์ตัน [10] ที่นำเสนอ Practical Approach สำหรับสกัดข้อความออกจากเอกสารต้นฉบับเพื่อย่อความ โดยแบ่งกลุ่มคำสำคัญ ค้นหาความสัมพันธ์ของ Text Span ใช้วิธีการ Local Properties สำหรับแบ่งกลุ่มคำสำคัญแต่ละย่อหน้า และใช้วิธีการ Global สำหรับพิจารณาความสัมพันธ์ของทุกย่อหน้า แต่งานวิจัยนี้ยังมีข้อจำกัดหากเอกสารมีความยาวน้อยเกินไป อาจทำให้ผลลัพธ์ในการย่อความผิดพลาด และผลลัพธ์ในการย่อความยังพบคำที่ไม่จำเป็นรวมอยู่ด้วย ส่วนการย่อความเอกสารภาษาไทย โดยการวิเคราะห์หาความหมายที่แอบแฝง (Latent Semantic) [11] นำมาช่วยในการย่อความได้ดี แต่ยังมีข้อจำกัดในการหาขอบเขตคำ วลี และการตัดคำ ซึ่งผู้วิจัยได้ให้ข้อเสนอแนะว่าควรนำโมเดลทางภาษาศาสตร์มาใช้ร่วมกับการประมวลผลของภาษาธรรมชาติ เพื่อช่วยลดข้อผิดพลาดในการตัดคำ และเพิ่มประสิทธิภาพในการย่อความมากยิ่งขึ้น

## 2.2 การวิเคราะห์องค์ประกอบ (Factor Analysis)

การวิเคราะห์องค์ประกอบ เป็นเทคนิคที่ช่วยในการจับกลุ่มรวมกลุ่ม หรือรวมตัวแปรที่มีความสัมพันธ์กันไว้ในกลุ่มหรือปัจจัยเดียวกัน [12] ซึ่งการวิเคราะห์องค์ประกอบถูกออกแบบมา เพื่อใช้ในการตรวจสอบโครงสร้างของชุดตัวแปรและอธิบายความสัมพันธ์ระหว่างตัวแปร ในรูปของจำนวนที่น้อยที่สุดของตัวแปรแฝงที่สังเกตไม่ได้ ซึ่งตัวแปรแฝงที่สังเกตไม่ได้เหล่านี้จะถูกเรียกว่า “องค์ประกอบ”

### 2.2.1 ประโยชน์ของการวิเคราะห์องค์ประกอบ

การวิเคราะห์องค์ประกอบ จะช่วยลดจำนวนตัวแปรโดยการรวมตัวแปรให้อยู่ในปัจจัยเดียวกัน ซึ่งปัจจัยที่ได้ก็คือเป็นตัวแปรใหม่ ที่สามารถหาค่าข้อมูลของปัจจัยที่สร้างขึ้นได้ เรียกว่า “ค่าคะแนนปัจจัย (Factor Score)” จึงสามารถนำปัจจัยดังกล่าวไปเป็นตัวแปร สำหรับการวิเคราะห์ทางสถิติต่อไป และยังช่วยในการแก้ปัญหา อันเนื่องมาจากการที่ตัวแปรอิสระของเทคนิคการวิเคราะห์สมการความถดถอย มีความสัมพันธ์กัน (Multicollinearity) ซึ่งสามารถแก้ปัญหาได้โดยทำการรวมตัวแปรอิสระที่มีความสัมพันธ์ไว้ด้วยกัน สร้างตัวแปรใหม่แล้วนำไปเป็นตัวแปรอิสระในการวิเคราะห์ความถดถอยต่อไป

และการวิเคราะห์องค์ประกอบ ซึ่งสามารถแสดงให้เห็นถึงโครงสร้างความสัมพันธ์ของตัวแปรที่ศึกษา จึงสามารถวิเคราะห์โครงสร้างที่แสดงความสัมพันธ์ของตัวแปรต่างๆ ที่อยู่ในปัจจัยเดียวกันได้ ทำให้สามารถอธิบายความหมายของแต่ละปัจจัยได้

2.2.2 วัตถุประสงค์ของการวิเคราะห์องค์ประกอบสามารถแบ่งได้เป็น 2 แนวทางดังนี้

ก) การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploration Factor Analysis: EFA) เป็นโมเดลสำหรับศึกษาตัวประกอบร่วมหรือค้นหาตัวแปรแฝงที่ซ่อนอยู่ภายใต้ตัวแปรที่สังเกตหรือวัดได้ ซึ่งการวิเคราะห์องค์ประกอบจะใช้ในการสำรวจและระบุองค์ประกอบร่วมที่สามารถอธิบายความสัมพันธ์ระหว่างตัวแปร ผลจากการวิเคราะห์องค์ประกอบจะช่วยให้นักวิจัยลดจำนวนตัวแปรลง และได้องค์ประกอบซึ่งทำให้เข้าใจลักษณะของข้อมูลได้ง่าย สะดวกในการแปลความหมาย รวมทั้งได้ทราบแบบแผน (Pattern) และโครงสร้าง (Structure) ความสัมพันธ์ของข้อมูลด้วย [13]

ข) การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) เป็นโมเดลที่ใช้สำหรับพิสูจน์ตรวจสอบ หรือยืนยันทฤษฎีที่ผู้เฝ้าค้นพบว่าตรงกับโมเดลหรือทฤษฎีที่มีอยู่หรือไม่ โดยทำการทดสอบสมมติฐานเกี่ยวกับแบบแผน และโครงสร้างความสัมพันธ์ของข้อมูลของผู้วิจัยสนใจ

### 2.3 ขั้นตอนการวิเคราะห์องค์ประกอบ

มีรายละเอียดดังนี้

2.3.1 การสร้างเมตริกซ์ความสัมพันธ์ระหว่างคู่ของตัวแปรทุกตัว (Correlation Matrix)

โดยทำการสร้างเมตริกซ์ความสัมพันธ์ ระหว่างคู่ของตัวแปรทุกตัว เพื่อวิเคราะห์ปัจจัยที่จะหาความสัมพันธ์ในรูปแบบเส้นตรงระหว่างตัวแปรทุกคู่ที่ต้องการจัดกลุ่มให้อยู่ในรูปแบบ Correlation Matrix โดยทำการสร้าง และตรวจสอบความสัมพันธ์ระหว่างตัวแปรที่สนใจ ว่ามีความสัมพันธ์กันหรือไม่

2.3.2 สกัดองค์ประกอบของตัวแปร (Factor Extraction)

เพื่อทำการหาองค์ประกอบว่าจะสามารถใช้ตัวแปรได้ทุกตัวหรือไม่ ซึ่งวิธีที่นิยมใช้ได้แก่

ก) วิธีการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis: PCA) โดยมีหลักการว่าองค์ประกอบแต่ละตัวจะสามารถเขียนเป็น Linear Combination ของตัวแปรอะไรก็ได้บ้าง ซึ่งจะเป็นการคำนวณหาค่าน้ำหนักของตัวแปรย่อย เพื่อสร้างเป็นตัวแปรองค์ประกอบใหม่

ข) วิธีการวิเคราะห์องค์ประกอบร่วม (Common Factor

Analysis: CFA) ซึ่งมีหลายวิธีการ เช่น วิธีกำลังสองน้อยที่สุดที่ไม่ปรับน้ำหนัก (Unweighted Least Square) เป็นวิธีการสกัดปัจจัยโดยกำหนดจำนวนไว้ตายตัว และพยายามหาเมตริกซ์แบบแผนของปัจจัย (Factor Pattern Matrix) ที่ทำให้ผลรวมของความแตกต่างกำลังสองระหว่างเมตริกซ์ที่คำนวณได้ หรือเมตริกซ์ความสัมพันธ์ที่สร้างขึ้นใหม่ และเมตริกซ์ความสัมพันธ์เดิมระหว่างตัวแปรมีค่าน้อยที่สุด วิธีกำลังสองน้อยที่สุดทั่วไป (Generalized Least Square: GLS) เป็นวิธีการที่ใช้การถ่วงน้ำหนักความสัมพันธ์ระหว่างตัวแปร ในเชิงปฏิภาคกลับ กับความเด่นเฉพาะ (Uniqueness) ของตัวแปรนั้น วิธีความเป็นไปได้สูงสุด (Maximum Likelihood: ML) เป็นวิธีการที่ใช้หาค่าโดยการประมาณที่สามารถจะใช้เมตริกซ์ความสัมพันธ์ที่คำนวณได้ใกล้กับเมตริกซ์ ที่ได้จากการสังเกตโดยสมมติว่าข้อมูลนั้นเป็นข้อมูลตัวอย่างที่มีการกระจายปกติหลายตัวแปร (Multivariate Normal Distribution) วิธีอัลฟา (Alpha Method) เป็นวิธีการที่มีการตั้งข้อสมมติฐานไว้ว่าตัวแปรแต่ละตัวมีส่วนประกอบ 2 ส่วน คือ ปัจจัยร่วม และปัจจัยเฉพาะ แต่ที่แตกต่างจากวิธีการอื่นๆ คือ แทนที่จะถือว่าจำนวนกรณีที่จะใช้ในการวิเคราะห์เป็นจำนวนตัวอย่าง แต่ถือว่าจำนวนตัวแปรนั้นเป็นตัวอย่างเป็นคุณสมบัติของประชากร ซึ่งวัตถุประสงค์ของวิธีการแยกปัจจัยโดยอาศัยค่าอัลฟา คือ หาปัจจัยที่เป็นตัวแทนของคุณสมบัติของประชากร หรือมีอัตราการใช้ได้ทั่วไปสูงสุด (Maximum Generalizability) และวิธีเงา (Image Method) เป็นวิธีการแยกปัจจัยอีกวิธีหนึ่ง ซึ่งสมมติ ว่าตัวแปรแต่ละตัวแบ่งออกได้เป็น 2 ส่วน คือ ส่วนที่เกิดจาก ปัจจัยร่วมและส่วนที่เกิดจากปัจจัยเฉพาะสัดส่วนที่แน่นอนของทั้งสองส่วน ซึ่งคำนวณได้จากการประมาณ โดยอาศัยเมตริกซ์ความสัมพันธ์ระหว่างตัวแปรทฤษฎีเงา (Image Theory)

2.3.3 การหมุนแกนองค์ประกอบ (Factor Rotation)

การหมุนแกนองค์ประกอบ เป็นขั้นตอนการแยกตัวแปรย่อยที่ศึกษาควรจัดอยู่ในกลุ่ม หรือในองค์ประกอบใด ซึ่งการหมุนแกนจะมีอยู่ 2 วิธี คือ

ก) การหมุนแกนแบบตั้งฉาก (Orthogonal Rotation) เพื่อดูว่าตัวแปรย่อยได้อยู่ใกล้แกนขององค์ประกอบ ก็ให้จัดตัวแปรย่อยนั้นอยู่ในองค์ประกอบดังกล่าว เป็นการหมุนแกนแบบที่ปัจจัยแต่ละปัจจัยไม่มีความสัมพันธ์กัน

ข) การหมุนแกนแบบมุมแหลม (Oblique Rotation) เป็นการหมุนแกนแบบให้แกนของปัจจัย หมุนจากตำแหน่งเดิมในลักษณะมุมแหลม และไม่ตั้งฉากกันตลอดเวลาที่มีการหมุนแกน



โดยการหมุนแกนแบบนี้จะระบุระดับความสัมพันธ์ระหว่างปัจจัย โดยกำหนดจำนวนองศาของมุมแหลมตั้งแต่ 0 ถึง 90 องศา ถ้าต้องการให้ปัจจัยที่ได้มีความสัมพันธ์กันสูงให้กำหนดค่าจำนวนองศาให้ต่ำ และถ้าต้องการให้ปัจจัยที่ได้สัมพันธ์กันน้อยให้กำหนดค่าจำนวนองศาให้สูง

### 3. วิธีดำเนินการวิจัย

งานวิจัยนี้เป็นการศึกษาเชิงสำรวจและการวิจัยเชิงยืนยันที่ดำเนินงานอย่างมีลำดับขั้นตอนดังนี้

ขั้นที่ 1 ทำการศึกษาถึงปัจจัยที่เกี่ยวข้องกับการย่อความข่าวภาษาไทย โดยศึกษาถึงรูปแบบและวิธีการสร้างแบบสอบถาม

ขั้นที่ 2 กำหนดประเด็นคำถาม และร่างข้อคำถามให้ครอบคลุมทุกประเด็นที่เกี่ยวข้องกับการย่อความ

ขั้นที่ 3 จัดทำแบบสอบถามฉบับร่าง

ขั้นที่ 4 เสนออาจารย์ที่ปรึกษาวิทยานิพนธ์และผู้เชี่ยวชาญพิจารณาเพื่อปรับปรุงแก้ไข

ขั้นที่ 5 นำแบบสอบถามฉบับร่างที่แก้ไขแล้วไปให้กลุ่มตัวอย่างทดลองใช้ เพื่อหาความเชื่อมั่นของแบบสอบถาม

ขั้นที่ 6 ปรับปรุงแบบสอบถาม และจัดทำแบบสอบถามฉบับสมบูรณ์ เพื่อเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่าง

ขั้นที่ 7 ทำการเก็บรวบรวมและวิเคราะห์ข้อมูล

ขั้นที่ 8 ศึกษาองค์ประกอบที่เหมาะสม และมีความสัมพันธ์ต่อการย่อความ และวิเคราะห์องค์ประกอบเชิงยืนยันโดยใช้โปรแกรม LISREL

#### 3.1 กลุ่มตัวอย่างที่ใช้ในการวิจัย

กลุ่มตัวอย่างที่ใช้ในการวิจัยแบ่งได้เป็น 2 กลุ่ม ได้แก่ กลุ่มแรกสำหรับหาความเชื่อมั่นของแบบสอบถาม ซึ่งเป็นครูสอนนิสิตภาษาไทย จำนวน 30 คน และกลุ่มที่สองเป็นผู้เชี่ยวชาญสำหรับตอบแบบสอบถาม ซึ่งเป็นครูภาษาไทยที่มีประสบการณ์ในการย่อความ จำนวน 96 คน

#### 3.2 เครื่องมือที่ใช้ในการวิจัยและการสร้างแบบสอบถาม

ผู้วิจัยใช้แบบสอบถามเป็นเครื่องมือในการวิจัย ซึ่งเป็นแบบสอบถามที่มีมาตราส่วนประมาณค่า 5 ระดับ ที่แบ่งออกเป็น 3 ตอน คือ ตอนที่ 1 ข้อมูลส่วนตัวของผู้ตอบแบบสอบถาม ตอนที่ 2 ขั้นตอนวิธีการย่อความข่าวและความคิดเห็นของตอบแบบสอบถาม เกี่ยวกับปัจจัยที่มีผลต่อการย่อความข่าว ซึ่งสองตอนแรกเป็นคำถามปลายปิด (Closed Form) ที่สอบถามเพื่อต้องการคำตอบเฉพาะเจาะจง และตอนที่ 3 ข้อเสนอแนะต่าง ๆ ในแต่ละด้าน เป็นคำถามปลายเปิด (Opened Form) ที่

Position	School	Level_Ed	Experience	Task	Exp_Sum	Year_Ex	ID01
1	ครูชำนาญการ	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย ม. 1 และ 3	มี	10 ปีขึ้นไป	4
2	ครู ก. 3	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย ม. 4	มี	10 ปีขึ้นไป	4
3	ครู	ครูประถมชว	15 ปีขึ้นไป	รอดักต์หลักสูตรการศึกษา	มี	10 ปีขึ้นไป	3
4	ครูชำนาญการ	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย	มี	10 ปีขึ้นไป	2
5	ครู	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย	มี	10 ปีขึ้นไป	1
6	ครู	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย ม. 3	มี	10 ปีขึ้นไป	4
7	ครู ก. 3	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย วรรณคดีเจ้าฟ้า ม...	มี	10 ปีขึ้นไป	5
8	...	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย	มี	10 ปีขึ้นไป	3
9	ครู ก. 2	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย ม. 3	มี	10 ปีขึ้นไป	3
10	ครู ก. 3	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย	มี	10 ปีขึ้นไป	4
11	ครู ก. 3	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย ม. 6	มี	10 ปีขึ้นไป	4
12	ครู ก. 3	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย ม. 5	มี	10 ปีขึ้นไป	4
13	ครู	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย	มี	10 ปีขึ้นไป	3
14	ครู	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย ม. 1 15 คาบต่อสัปดาห์	มี	10 ปีขึ้นไป	5
15	ครู	ครูประถมชว	15 ปีขึ้นไป	เป็นนักเรียนพิเศษ สอนสอนภาษาไทย	มี	10 ปีขึ้นไป	2
16	ครู ก. 3	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย ม. 1 15 คาบต่อสัปดาห์	มี	10 ปีขึ้นไป	4
17	ครู ก. 2	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย ม. 3	มี	10 ปีขึ้นไป	3
18	ครู	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย	มี	10 ปีขึ้นไป	3
19	ครู	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย ม. 1	มี	7 ปีขึ้นไป	5
20	ครู ก. 2	ครูประถมชว	15 ปีขึ้นไป	สอนภาษาไทย ม. 4 และ 5	มี	10 ปีขึ้นไป	5
21	ครูชำนาญการ	ครูประถมชว	15 ปีขึ้นไป	หัวหน้างานบริหาร และสอนภาษาไทย ม.3	มี	10 ปีขึ้นไป	4
22	ครู ก. 3	ครูประถมชว	15 ปีขึ้นไป	หัวหน้างานบริหาร และสอนภาษาไทย	มี	10 ปีขึ้นไป	5

ภาพที่ 1 การวิเคราะห์ข้อมูลด้วยโปรแกรม SPSS for Windows

ไม่กำหนดคำตอบไว้อย่างเฉพาะเจาะจง เป็นการเปิดโอกาสให้ผู้ตอบตอบได้อย่างอิสระ [14]

หลังจากจัดทำแบบสอบถามฉบับร่าง และทำการปรับปรุงแก้ไขตามข้อเสนอนี้ของอาจารย์ที่ปรึกษาวิทยานิพนธ์ และผู้เชี่ยวชาญแล้ว จึงนำแบบสอบถามดังกล่าวไปให้กลุ่มตัวอย่างที่เป็นครูภาษาไทยจำนวน 30 คน ตอบแบบสอบถามเพื่อหาความเชื่อมั่นของเครื่องมือวัด โดยภาพรวมได้ค่า Cronbach's Alpha มีค่าเท่ากับ 0.803 ซึ่งเข้าใกล้ 1 แสดงว่าแบบสอบถามที่สร้างขึ้นมีความน่าเชื่อถือสูง

#### 3.3 การเก็บรวบรวมและวิเคราะห์ผลข้อมูล

เมื่อได้แบบสอบถามฉบับสมบูรณ์แล้ว ผู้วิจัยจึงนำแบบสอบถามดังกล่าวไปเก็บรวบรวมข้อมูลจากผู้เชี่ยวชาญที่เป็นครูภาษาไทย ซึ่งมีประสบการณ์ในการย่อความ จำนวน 96 คน เมื่อเก็บรวบรวมข้อมูลทั้งหมดแล้ว ได้นำมาตรวจสอบความถูกต้องของข้อมูล โดยทำการกำหนดชุดตัวแปร และรหัสข้อมูล เพื่อป้อนข้อมูลลงในโปรแกรม SPSS for Windows

### 4. ผลการวิจัย

จากการวิเคราะห์ข้อมูลด้วยโปรแกรม SPSS for Windows แสดงให้เห็นถึงข้อมูลส่วนตัวของผู้เชี่ยวชาญ ซึ่งมีรายละเอียดดังตารางที่ 1

จากแบบสอบถามตอนที่ 2 ขั้นตอนวิธีการย่อความข่าว และความคิดเห็นของผู้เชี่ยวชาญเกี่ยวกับปัจจัยที่มีผลต่อการย่อความข่าวภาษาไทยที่ใช้กรณีศึกษาเป็นข่าววิทยุการ และข่าวเศรษฐกิจ โดยเน้นการวิเคราะห์โครงสร้างเนื้อหาของเอกสารต้นฉบับ เพื่อนำปัจจัยดังกล่าวไปใช้พัฒนาระบบย่อความเอกสารภาษาไทยให้มีประสิทธิภาพต่อไป โดยผลการวิจัยของแบบ



ตารางที่ 1 ข้อมูลส่วนตัวของผู้ตอบแบบสอบถาม

ข้อมูลของผู้ตอบแบบสอบถาม	ความถี่	ร้อยละ
ระดับการศึกษา (N=96)		
ปริญญาตรี	87	90.6
ปริญญาโท	7	7.3
ไม่ได้กรอกข้อมูล	2	97.9
ประสบการณ์ในการทำงาน (N=96)		
ต่ำกว่า 5 ปี	4	4.2
5 ปีแต่ไม่ถึง 10 ปี	5	5.2
10 ปีแต่ไม่ถึง 15 ปี	1	1
15 ปีขึ้นไป	83	86.5
ไม่ได้กรอกข้อมูล	3	3.1
ประสบการณ์ในการย่อความ (N=96)		
ต่ำกว่า 1 ปี	3	3.1
1 ปีแต่ไม่ถึง 3 ปี	4	4.2
4 ปีแต่ไม่ถึง 6 ปี	2	2.1
7 ปีแต่ไม่ถึง 10 ปี	11	11.5
10 ปีขึ้นไป	71	74
ไม่ได้กรอกข้อมูล	5	5.2

สอบถามตอนที่ 2 มีรายละเอียดดังนี้

ด้านส่วนนำเข้าข้อมูล ประกอบด้วยปัจจัยต่าง ๆ ได้แก่ โครงสร้างของเอกสารต้นฉบับ ขนาดของเอกสารต้นฉบับ ความยาวของเอกสารต้นฉบับ รูปแบบของภาษาที่ใช้ในการบรรยายเอกสารต้นฉบับ และลักษณะการใช้งานเอกสารต้นฉบับ ซึ่งผู้เชี่ยวชาญมีระดับความคิดเห็นในด้านนี้อยู่ในระดับมาก ( $\bar{X} = 3.67, S.D. = 1.05$ )

ด้านกระบวนการย่อความ ประกอบด้วยปัจจัยต่าง ๆ ได้แก่ โครงสร้างและหลักภาษาศาสตร์ที่ปรากฏในเอกสารต้นฉบับ สถานการณ์ที่เกิดขึ้นในเอกสารต้นฉบับ ผู้กระทำ หรือผู้ที่เกี่ยวข้องในเอกสารต้นฉบับ เป้าหมายในการย่อความ จำนวนคำที่ปรากฏในประโยค คำสำคัญที่ปรากฏในเอกสารต้นฉบับ ข้อเท็จจริงที่ปรากฏอยู่ในเอกสารต้นฉบับ และเทคนิคการเลือกประโยคสำคัญในการย่อความ ซึ่งผู้เชี่ยวชาญมีระดับความคิดเห็นในด้านนี้อยู่ในระดับมาก ( $\bar{X} = 4.34, S.D. = 0.76$ )

ด้านผลลัพธ์ที่ได้จากการย่อความ ประกอบด้วยปัจจัยต่าง ๆ ได้แก่ ขนาดของผลลัพธ์ หัวข้อเรื่องของการย่อความ คำหยุด คำที่ไม่ควรปรากฏในผลลัพธ์ รูปประโยคของผลลัพธ์ และลักษณะของการแสดงผลการย่อความ ซึ่งผู้เชี่ยวชาญมีระดับความคิดเห็นในด้านนี้อยู่ในระดับมาก ( $\bar{X} = 3.94, S.D. = 1.05$ )

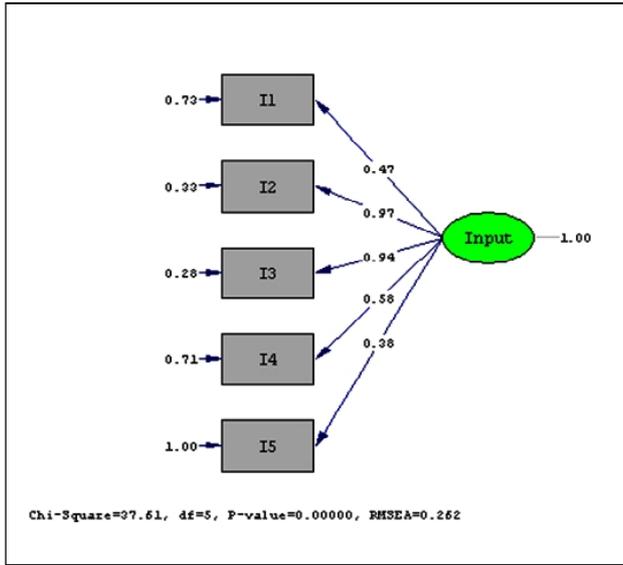
ดังนั้นโดยภาพรวมของทุกด้านผู้เชี่ยวชาญมีระดับความคิดเห็นอยู่ในระดับมาก ( $\bar{X} = 4.03, S.D. = 0.99$ ) แสดงว่าปัจจัยที่ได้ทำการศึกษามีผลต่อประสิทธิภาพของการย่อความข่าวภาษาไทย

#### 4.2 ผลการวิเคราะห์ปัจจัยเชิงยืนยัน

จากนั้นทำการวิเคราะห์ข้อมูลโดยใช้โปรแกรม LISREL เพื่อวิเคราะห์ปัจจัยเชิงยืนยัน (Confirmatory Factor Analysis: CFA) ที่มีผลต่อการย่อความข่าวภาษาไทย ซึ่งเป็นการวิเคราะห์องค์ประกอบสำหรับพิสูจน์ ตรวจสอบ หรือยืนยันทฤษฎีที่ค้นพบว่าตรงกับโมเดลหรือทฤษฎีที่มีอยู่หรือไม่ โดยวิเคราะห์ค่าน้ำหนักองค์ประกอบ (Factors Loading) เมื่อได้ผลการวิเคราะห์จะทำการตรวจสอบความกลมกลืนระหว่างโมเดลตามทฤษฎีกับข้อมูลเชิงประจักษ์ ตรวจสอบโครงสร้างของโมเดลว่ามีความแตกต่างระหว่างกลุ่มตัวอย่างหลายกลุ่มหรือไม่ ซึ่งจะทำให้การระบุตัวแปรที่ต้องการนำเข้าสู่กระบวนการวิเคราะห์องค์ประกอบ โดยค่าสถิติที่นำมาใช้คือ KMO (Kaiser Meyer Olkin) and Barlett's Test of Sphericity เพื่อวัดความพอเพียงของกลุ่มตัวแปรที่นำมาใช้งาน

จากการหาค่า KMO ได้ค่าสัมประสิทธิ์ KMO เท่ากับ 0.723 ซึ่งมากกว่า 0.5 แสดงว่าจำนวนกลุ่มตัวอย่างที่ใช้พอเพียงในการทำวิจัย จากนั้นนำปัจจัยที่ได้จากการวิเคราะห์องค์ประกอบเชิงสำรวจมาสร้างโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยัน โดยจะต้องพิจารณาค่า P-value ให้มีค่ามากกว่าหรือเท่ากับ 0.05 และพิจารณาค่าไค-สแควร์ (Chi-Square Statistics) ให้มีค่าเข้าใกล้ศูนย์ เนื่องจากยิ่งถ้ามีค่าเข้าใกล้ศูนย์ แสดงว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ โดยทำการวิเคราะห์ปัจจัยที่ส่งผลการย่อความข่าวภาษาไทยดังที่กล่าวไปแล้ว เพื่อทำการตรวจสอบ และยืนยันถึงปัจจัยที่ได้ศึกษาว่าส่งผลการย่อความข่าวภาษาไทย โดยคำนวณหาค่าน้ำหนักองค์ประกอบ และตรวจสอบความกลมกลืนระหว่างโมเดลตามทฤษฎีกับข้อมูลเชิงประจักษ์ จากนั้นทำการวิเคราะห์ปัจจัยโดยใช้วิธีการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis: PCA) เพื่อคำนวณหาค่าน้ำหนักของตัวแปรแบบ PCA และทำการหมุนแกนแบบ Varimax เพื่อปรับค่าข้อมูลให้มีความเหมาะสม และนำปัจจัยที่ได้ศึกษามาสร้างโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันในแต่ละด้าน ตัวอย่างปัจจัยด้านส่วนนำเข้าข้อมูลก่อนทำการปรับค่าสามารถแสดงได้ดังภาพที่ 2

โดยปัจจัยด้านส่วนนำเข้าข้อมูลมีค่า P-value = 0.00000



ภาพที่ 2 ปัจจัยด้านส่วนนำเข้าข้อมูลที่มีผลต่อการย่อความก่อนทำการปรับค่า

และมีค่าไค-สแควร์ = 37.61 จึงต้องปรับค่าข้อมูลให้มีความเหมาะสม โดยตัวอย่างคำสั่ง SQL ของโปรแกรม LISREL ในการวิเคราะห์หองค์ประกอบด้านนำเข้าข้อมูล สามารถแสดงได้ดังภาพที่ 3 ซึ่งประกอบด้วยปัจจัยหรือองค์ประกอบ คือ โครงสร้างของเอกสารต้นฉบับ (I1) ขนาดของเอกสารต้นฉบับ (I2) ความยาวของเอกสารต้นฉบับ (I3) รูปแบบของภาษาที่ใช้ในการบรรยายเอกสารต้นฉบับ (I4) และลักษณะการใช้งานเอกสารต้นฉบับ (I5)

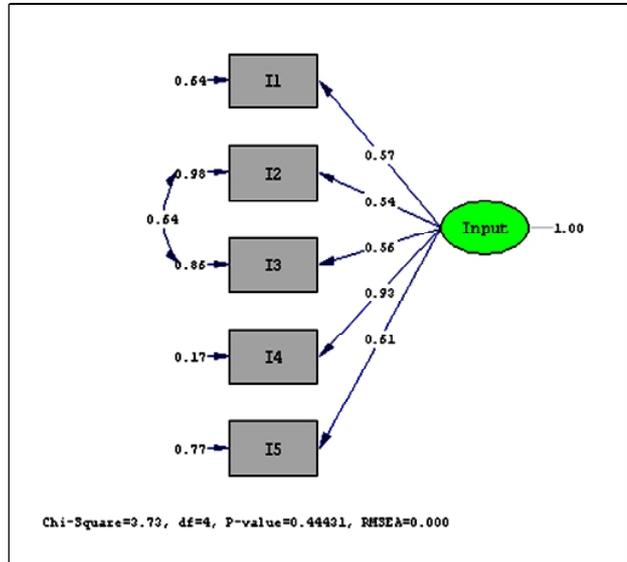
```

TEXT SUMMARIZATION MEASUREMENT MODEL
DA NI=5 NO=96 NG=1 MA=CM
LA
I1 I2 I3 I4 I5
KM
1
.391      1
.371      .772      1
.526      .424      .465      1
.273      .260      .256      .533      1
SD
.978      1.129      1.081      1.021      1.069
MO NX=5 NK=1 LX=FU,FI PH=DI,FR TD=FU,FI
FR LX(1,1) LX(2,1) LX(3,1) LX(4,1) LX(5,1)
FR TD(1,1) TD(2,2) TD(3,3) TD(4,4) TD(5,5) TD(3,2)
LK
Input
PD
OU SE TV EF RS SS FS MI ND=3
    
```

ภาพที่ 3 คำสั่ง SQL ของโปรแกรม LISREL

จากการปรับค่าองค์ประกอบให้มีความเหมาะสม โดยหมุนแกนแบบ Varimax ซึ่งสามารถหมุนแกนได้ไม่เกิน 5 รอบ (ค่า df = 5) ในการปรับค่าได้หมุนแกน จำนวน 1 รอบ จึง

ได้ค่า P-value = 0.44431 และค่าไค-สแควร์ = 3.73 ซึ่งทำให้ค่า P-value มีค่ามากกว่า หรือเท่ากับ 0.05 และค่าไค-สแควร์ มีค่าเข้าใกล้ศูนย์ ดังแสดงในภาพที่ 4 แสดงว่าโมเดลของการย่อความด้านส่วนนำเข้าข้อมูลที่มีผลต่อการย่อความมีความเหมาะสม และสอดคล้องกับโมเดลเชิงประจักษ์



ภาพที่ 4 ปัจจัยด้านส่วนนำเข้าข้อมูลที่มีผลต่อการย่อความหลังทำการปรับค่า

ปัจจัยด้านกระบวนการย่อความที่ทำการวิเคราะห์องค์ประกอบเชิงยืนยันด้วยโปรแกรม LISREL ก่อนหมุนแกนแบบ Varimax เพื่อปรับค่าข้อมูล มีค่า P-value = 0.00000 และมีค่าไค-สแควร์ = 69.49 จึงต้องทำการปรับค่าตัวแปรให้มีความเหมาะสม โดยทำหมุนแกน จำนวน 5 รอบ จึงได้ค่า P-value = 0.39438 และค่าไค-สแควร์ = 15.82 ซึ่งแสดงได้ว่าโมเดลของการย่อความด้านกระบวนการที่มีผลต่อการย่อความมีความเหมาะสม และมีความสอดคล้องกับโมเดลเชิงประจักษ์

ปัจจัยด้านผลลัพธ์ที่ได้จากการย่อความที่ทำการวิเคราะห์องค์ประกอบเชิงยืนยันด้วยโปรแกรม LISREL ก่อนหมุนแกนแบบ Varimax เพื่อปรับค่าข้อมูลให้เหมาะสม มีค่า P-value = 0.07760 และมีค่าไค-สแควร์ = 22.05 จึงทำการหมุนแกนเพื่อลดค่าไค-สแควร์ให้เข้า 0 โดยทำการหมุนแกนจำนวน 6 ครั้ง จึงได้ค่า P-value = 0.89062 และค่าไค-สแควร์ = 3.61 ซึ่งแสดงได้ว่าโมเดลด้านผลลัพธ์ที่ได้จากการย่อความมีความเหมาะสม และมีความสอดคล้องกับโมเดลเชิงประจักษ์

5. สรุปผลและข้อเสนอแนะ

ในการพัฒนาระบบย่อความเอกสารภาษาไทยนั้น จะต้องคำนึงถึงโครงสร้างของภาษาไทย ที่มีปัญหาในการประมวลผล



ในระดับคำ และระดับวลี ผู้วิจัยจึงต้องทำการศึกษา และวิเคราะห์ถึงปัจจัยต่างๆ ที่มีผลต่อการย่อความเอกสารภาษาไทยจากผลการวิจัยที่ใช้กรณีศึกษาเป็นข่าว และทำการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างที่เป็นครูภาษาไทย ที่มีประสบการณ์ในการย่อความ จำนวน 96 คน โดยเน้นวิเคราะห์จากโครงสร้างของเนื้อหาเอกสารต้นฉบับ พบว่าปัจจัยที่ส่งผลต่อการย่อความข่าวภาษาไทย แบ่งได้เป็น 3 ด้าน คือ ด้านส่วนนำเข้าข้อมูล ( $\bar{x} = 3.67$ , S.D. = 1.05) ด้านกระบวนการย่อความ ( $\bar{x} = 4.34$ , S.D. = 0.76) และด้านผลลัพธ์ที่ได้จากการย่อความ ( $\bar{x} = 3.94$ , S.D. = 1.05) ซึ่งโดยภาพรวมผู้เชี่ยวชาญมีระดับความคิดเห็นอยู่ในระดับมาก ( $\bar{x} = 4.03$ , S.D. = 0.99) แสดงว่าปัจจัยที่ได้ทำการศึกษามีผลต่อการย่อความข่าวภาษาไทย และผลการวิเคราะห์หองค์ประกอบตามโมเดลการวัดองค์ประกอบเชิงสัมพันธ์ที่มีผลต่อการย่อความข่าว พบว่าหลังจากปรับค่าข้อมูลให้มีความเหมาะสม ทุกด้านที่ทำการศึกษามีค่า P-value มากกว่า 0.05 และค่าไค-สแควร์ (Chi-Square Statistics) มีค่าเข้าใกล้ศูนย์ แสดงว่าแต่ละด้านมีความสอดคล้องกับข้อมูลเชิงประจักษ์ที่ ซึ่งแตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติที่ระดับ 0.05 ซึ่งปัจจัยต่าง ๆ เหล่านี้ ผู้พัฒนาระบบย่อความเอกสารภาษาไทยจะต้องให้ความสำคัญ และคำนึงถึงเพื่อสามารถพัฒนาระบบได้อย่างอัตโนมัติ ตรงกับความต้องการของผู้ใช้งาน และสามารถรองรับการย่อความได้อย่างมีประสิทธิภาพ ซึ่งในการพัฒนาระบบย่อความเอกสารนั้นไม่ควรจำกัดอยู่ในขอบเขตงานหนึ่ง ๆ จึงจะสามารถนำไปประยุกต์ใช้งานได้กว้างขวาง

### เอกสารอ้างอิง

- [1] ชัชวาล สังคีตตระการ และคนอื่น ๆ. "อับดุล : ระบบบริการข้อมูลเชิงบูรณาการผ่านระบบสนทนาออนไลน์". วารสารวิชาการเนคเทค, 8 (กรกฎาคม – ตุลาคม 2551): 318-323.
- [2] ชุติรัตน์ จรัสกุลชัย และชนาตัย กริ่งไกร. การพิสูจน์หาหัวเรื่องที่มีนัยสำคัญ และการย่อความของข้อความภาษาไทยแบบอัตโนมัติ. ภาควิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยเกษตรศาสตร์, 2545.
- [3] ธนุศักดิ์ ธีบุญศิริ. การออกแบบ และพัฒนาระบบต้นแบบสำหรับการสืบค้นข้อมูลสารสนเทศภาษาไทย ด้วยดัชนีหลายหลายระดับ. วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์, 2543.
- [4] ผุสดี บุญรอด และอนิราช มิ่งขวัญ. "การวิเคราะห์โครงสร้างเนื้อหาของเอกสารต้นฉบับเพื่อย่อความภาษาไทย". In *Proceedings of 2<sup>nd</sup> National Conference on Computing and Information Technology, NCCIT2006*, Bangkok, Thailand, May 19-20, 2006.
- [5] สมหมาย เดชสุภา. ภาษาไทยธุรกิจ. กรุงเทพฯ: วิทยาเขตพันธชยการพระนคร, 2533.
- [6] Sparck, J. *Advances in Automatic Text Summarizing*. Cambridge MA : MIT Press, 1998.
- [7] Hovy E., and Lin C. "Automated Text Summarization in SUMMARIST". In *Proceedings of the ACL Workshop on Intelligence Text Summarization*, pp. 18-24, 1997.
- [8] Carbonell, J., and Goldstein, J. "The use of MMR, diversity based reranking for reordering documents and producing summaries. In *Proceedings of SIGIR*, 1998.
- [9] Massih R.A.et. al. "Automatic Text Summarization Based on Word-Clusters and Ranking Algorithms". *ECIR 2005, LNCS 3408*, pp.142–156, 2005
- [10] Jaruskulchai, C. and Kruengkrai, C. "A Practical Text Summarizer by Paragraph Extraction for Thai". *The Sixth International Workshop on Information Retrieval with Asian Languages (IRAL - 2003)*, July 7, 2003.
- [11] อัญญาณี แดงไทย. "การย่อความเอกสารภาษาไทยโดยการวิเคราะห์หาความหมายที่แอบแฝง". วิทยานิพนธ์วิทยาศาสตร์มหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ บัณฑิตวิทยาลัยมหาวิทยาลัยเกษตรศาสตร์, 2548.
- [12] ส้ารวน ชินจันทิก. การศึกษาปัจจัยเชิงสาเหตุที่มีอิทธิพลต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ของนักเรียนระดับมัธยมศึกษาตอนปลาย โรงเรียนสาธิตมหาวิทยาลัยขอนแก่น. ขอนแก่น : มหาวิทยาลัยขอนแก่น, 2547.
- [13] นงลักษณ์ วิรัชชัย. ความสัมพันธ์เชิงโครงสร้างเชิงเส้น (LISREL) สถิติวิเคราะห์สำหรับการวิจัยทางสังคมศาสตร์และพฤติกรรมศาสตร์. พิมพ์ครั้งที่ 2, กรุงเทพฯ: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, 2523.
- [14] ประกายรัตน์ สุวรรณ. คู่มือการใช้โปรแกรม SPSS เวอร์ชัน 12 สำหรับ Windows. กรุงเทพฯ: ซีเอ็ดดูเคชั่น, 2548.