

ระบบย่อความเอกสารเป็นระบบที่ช่วยในการหาวลีสำคัญที่ปรากฏอยู่ในเอกสารต้นฉบับ เพื่อย่อความเนื้อหาของเอกสารให้ผู้อ่านสามารถพิจารณาเนื้อหาสำคัญได้จากเพียงไม่กี่วลี ในการพัฒนาระบบย่อความภาษาไทยให้มีประสิทธิภาพนั้นทำได้ยาก เนื่องจากความซับซ้อนทางด้านภาษา อีกทั้งต้องมีการวิเคราะห์และตีความหมายเนื้อหา จึงอาจทำให้ได้ผลลัพธ์ที่ผิดพลาดหรือเข้าใจยาก ดังนั้นวิทยานิพนธ์นี้จึงมีวัตถุประสงค์เพื่อศึกษา ออกแบบขั้นตอนวิธี พัฒนาระบบ และประเมินประสิทธิภาพของระบบย่อความภาษาไทยโดยใช้กรณีศึกษา คือ ข่าววิทยากร และข่าวเศรษฐกิจ โดยผู้วิจัยได้ประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติร่วมกับฐานความรู้ออนโทโลยี และใช้กระบวนการย่อความแบบสกัดจากเอกสารต้นฉบับ (Extraction-Based Summarization) ที่ตัดข้อความให้สั้นลง เลือกเฟ้นข้อความถ้อยคำแล้วนำมาสรุปให้ได้สาระสำคัญ และใช้กระบวนการย่อความแบบสาระสังเขป (Abstraction Summarization) ที่เหมาะกับการย่อความข่าวหรือบทความวิชาการ ทำให้สามารถพิจารณาภาพรวมได้ทั้งหมด โดยระบบที่พัฒนาขึ้นนี้ผู้ใช้สามารถเลือกเอกสารที่ต้องการย่อความ กำหนดรายละเอียด และขนาดผลลัพธ์ที่ต้องการ จากนั้นระบบจะทำการย่อความเอกสารให้โดยอัตโนมัติ ทำให้เกิดความสะดวกรวดเร็วและลดเวลาในการย่อความเอกสาร

จากงานวิจัยได้ทำการศึกษาถึงปัจจัยที่ส่งผลต่อการย่อความ โดยวิเคราะห์ห่อองค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) ซึ่งใช้ค่าสถิติ KMO (Kaiser Meyer Olkin) และ Barlett's Test of Sphericity เพื่อวัดความพอเพียงของตัวอย่างที่ใช้ โดยกลุ่มตัวอย่างเป็นผู้เชี่ยวชาญที่มีประสบการณ์ในการย่อความภาษาไทย จำนวน 96 คน จากผลการทดลองพบว่าทุกด้านที่ศึกษาซึ่งได้แก่ ปัจจัยด้านส่วนนำเข้า (Input Factors) ปัจจัยด้านกระบวนการย่อความ (Process Factors) และปัจจัยด้านผลลัพธ์ที่ได้จากการย่อความ (Output Factors) มีค่า P-value มากกว่า .05 แสดงว่าปัจจัยแต่ละด้านที่ศึกษามีความเหมาะสม และสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ที่ระดับนัยสำคัญทางสถิติ .05 และจากการประเมินประสิทธิภาพของระบบพบว่า ระบบที่พัฒนาขึ้นมีประสิทธิภาพอยู่ในระดับมาก ($\bar{X} = 4.17$, S.D. = 0.78) และจากทดสอบความถูกต้องของระบบโดยนำผลลัพธ์ของการย่อความที่ได้จากระบบ เปรียบเทียบกับผลลัพธ์ที่ได้จากผู้เชี่ยวชาญโดยใช้เอกสารต้นฉบับชุดเดียวกันพบว่า ระบบที่พัฒนาขึ้นโดยภาพรวมมีค่าความถูกต้องเฉลี่ยในการเลือกประโยคที่สำคัญที่สุดเท่ากับ 50.01 เปอร์เซนต์ และผลลัพธ์ที่ได้จากการย่อความเอกสารมีค่าความถูกต้องเฉลี่ยเท่ากับ 70.08 เปอร์เซนต์ สรุปได้ว่าขั้นตอนวิธีและระบบที่พัฒนาขึ้นสามารถนำไปใช้ย่อความข่าวได้อย่างมีประสิทธิภาพ

Text summary system is the process of distilling the most important information from original documents into particular phrases which help readers getting the main idea in short forms. However, Thai text summary system is still in an early stage of developing mechanisms for automatic summarizing documents. It is difficult to develop text summary system to be efficient as a result of the complication of Thai text language which is needed to be analyzed and interpreted. Therefore, the objectives of this thesis are to study methodology, design summary algorithm, develop, and assess the effectiveness of Thai text summary system. The researcher does apply natural language processing with ontology-based knowledge. In addition, we have applied both extraction-based summarization techniques in order to shorten the context and to filter some statements and words, and abstraction summarization which are appropriate for academic journal summarization. Users can select the original document that he/she want to summarize. And then the system will automatically summarize the context which is easy and time saving.

In this research, we study key factors that affect the summarization. Confirmatory Factor Analysis (CFA), Kaiser Meyer Olkin (KMO) statistic and Barlett's Test of Sphericity are used to evaluate the sufficiency of samples in the research. The sample group consists of 96 Thai text precise experts. The result of the test indicates that all of the aspects which are studied, input factors, process factors and output factors are greater than .05 of P-value, so all analyzed factors correspond with empirical data very well. It is insignificant at the level of .05. The assessment of the system efficiency is good ($\bar{X} = 4.17$, S.D. = 0.78). The accuracy of the output from Thai text summarization system comparing to the output from Thai text precise experts shows that the average accuracy of the highlight sentence = 50.01% and the average accuracy of the summary output = 70.08%. We can conclude that the algorithm and the developed system can be used for news summarization efficiently.