

การเลือกประโยคสำคัญในการสรุปความภาษาไทย โดยใช้แบบจำลองแบบลำดับชั้น  
(Selection of Important Sentences in Thai Text Summarization Using a Hierarchical Model)

เพ็ญศิริ ลีตระกูล, เคนดวง ประดับสุวรรณ และ สุกรี สินธุภิญโญ

บทคัดย่อ

งานวิจัยด้านการสรุปความภาษาไทยยังมีจำนวนน้อย และส่วนใหญ่ใช้การคำนวณหาความสำคัญของประโยคโดยใช้การวิเคราะห์ความหมายแฝงหรือใช้การจัดกลุ่มประโยค ในงานวิจัยนี้ได้เสนอวิธีใหม่ในการเลือกประโยคที่สำคัญของเอกสารในการสรุปความภาษาไทย โดยการจัดเรียงประโยคเพื่อหาลำดับของประโยคที่มีระยะทางรวมสั้นที่สุด โดยใช้การค้นหาแบบเอสตาร์ (A\* Search) จัดเป็นชั้นตามระยะห่างจากประโยคซึ่งเป็นหัวข้อของบทความ จากนั้นแบ่งกลุ่มประโยคโดยพิจารณาจากคำที่ซ้ำกันและคำที่ไม่ซ้ำกัน แล้วจึงเลือกตัวแทนของกลุ่มประโยคเป็นประโยคสำคัญที่สามารถสรุปได้จากเอกสารภาษาไทย ชุดข้อมูลที่น่ามาทดสอบเป็นเอกสารภาษาไทยประเภทข่าวและบทความ 3 ประเภทคือ ข่าวเกษตร, ข่าวไอที-วิทยาการ และบทความไอที จำนวนทั้งหมด 30 เอกสาร ผลการทดลองพบว่า วิธีการที่นำเสนอให้ผลเลือกประโยคสำคัญที่ดีกว่าวิธีการวิเคราะห์หาความหมายแฝง

## **Abstract**

There have been a small number of Thai Text summarization studies. Most of them employ Latent Semantic Analysis or sentences clustering. This thesis proposes a new method in selection important sentence in Thai text summarization. Our approach starts with using A\* search algorithm to construct a shortest sentence sequence which is then organized as a hierarchical model sorted by distance from the title of a document. The sorted sentences are further clustered by considering redundant and non-redundant words. Finally, sentences representing each cluster are selected as the important sentences which can be summarizes from the document. To evaluate we tested our proposed technique with 30 documents collected from three categories of Thai news and articles - agricultural news, IT news, and IT articles. The results show that our method yields better selection results than the Latent Semantic Analysis method.