

ชื่อเรื่องการค้นคว้าแบบอิสระ

การสรุปเอกสารภาษาไทยที่มีการจัดโครงสร้างโดยใช้
เทคนิคการสรุปแฟร็กทัล

ผู้เขียน

นาย ชยพล แสนภิบาล

ปริญญา

วิทยาศาสตร์มหาบัณฑิต (วิทยาการคอมพิวเตอร์)

อาจารย์ที่ปรึกษาการค้นคว้าแบบอิสระ

ผศ. ดร. รัฐสิทธิ์ สุขะหุต

บทคัดย่อ

การค้นคว้าแบบอิสระนี้มีวัตถุประสงค์เพื่อนำเสนอวิธีการสรุปเอกสารภาษาไทยที่มีการจัดโครงสร้างโดยใช้เทคนิคการสรุปแฟร็กทัล โดยที่เทคนิคการสรุปแฟร็กทัลเป็นเทคนิคที่นำวิธีการทางสถิติมาใช้ในการคัดเลือกประโยคที่เป็นใจความสำคัญของเอกสารต้นฉบับโดยพิจารณาจากโครงสร้างลำดับชั้นและลักษณะเด่นของเอกสารต้นฉบับ ดังนั้นเอกสารต้นฉบับจะถูกจัดทำให้มีลักษณะเป็นโครงสร้างลำดับชั้นโดยใช้รูปแบบเลทเท็กซ์และนำคลังข้อมูลอรรถศาสตร์มาใช้ในการฝึกรอบรมชุดข้อมูลฝึกหัดเพื่อหาค่าน้ำหนักของคำต่าง ๆ ในชุดข้อมูลฝึกหัดเหล่านั้น ซึ่งค่าน้ำหนักของคำต่าง ๆ เหล่านี้จะมีค่าที่แตกต่างกันออกไปโดยพิจารณาจากตำแหน่งของคำที่ปรากฏอยู่ในเอกสารต้นฉบับ จากการทดลองพบว่าข้อสรุปที่ได้จากเทคนิคนี้มีความเที่ยงประมาณ 75% เมื่อเปรียบเทียบกับข้อสรุปที่ได้จากผู้เชี่ยวชาญ

Independent Study Title	Structured Thai Document Summarization Using Fractal Summarization Technique
Author	Mr. Chayapol Sanpiban
Degree	Master of Science (Computer Science)
Independent Study Advisor	Asst. Prof. Dr. Rattasit Sukhahuta

Abstract

The objective of this independent study is to propose the method of summarizing Thai document using fractal technique. Fractal summarization is a statistical method used for selecting the important sentences from salient feature of the source document. The input sentence must be in the structured format. We use LaTeX format to mark-up the structure of the document. Therefore, the document is organized into a hierarchical structure. In this study, we use ORCHID corpus in the training process as well as during testing of Thai document summarization. The weight for the words is computed according to their positions within the document. The experiment results showed that this technique yielded the results of 75% precision comparing to the summary results selected by the domain expertise.