

T 158868

ความก้าวหน้าทางเทคโนโลยีคอมพิวเตอร์ทำให้เอกสารอิเล็กทรอนิกส์มีปริมาณเพิ่มขึ้นอย่างรวดเร็ว เอกสารเหล่านี้ส่วนใหญ่เป็นภาษาอังกฤษ อย่างไรก็ตามเอกสารในภาษาอื่นก็มีเพิ่มมากขึ้นเป็นลำดับ ดังนั้นการบีบอัดข้อมูลเป็นวิธีหนึ่งในการเพิ่มประสิทธิภาพในการจัดเก็บข้อมูล

งานวิจัยนี้เสนอวิธีการบีบอัดข้อมูลที่มีประสิทธิภาพสูงสำหรับภาษาไทย โดยปรับปรุงวิธีการ Prediction by Partial Matching ซึ่งเป็นวิธีหนึ่งที่ทำให้ประสิทธิภาพการบีบอัดข้อมูลสูง โดยพิจารณาคูณลักษณะของภาษา

จากการทดลองเพื่อวัดประสิทธิภาพการบีบอัดข้อมูลของวิธีการที่นำเสนอกับข้อมูลลักษณะต่าง ๆ พบว่าวิธีการที่นำเสนอให้ประสิทธิภาพการบีบอัดข้อมูลได้ถึง 84 เปอร์เซ็นต์ ที่ขึ้นอยู่กับลักษณะของข้อมูลทดสอบ และรูปแบบการประมวลเบื้องต้น ซึ่งให้ผลที่ดีกว่ารหัสฮัฟแมนและสารสนเทศเฉลี่ย แต่อย่างไรก็ดีวิธีการในกลุ่ม Prediction by Partial Matching นี้เป็นที่ยอมรับกันว่าใช้เวลาในการเข้ารหัส-ถอดรหัสมากวิธีหนึ่ง ดังนั้นวิธีการที่นำเสนอเหมาะสมสำหรับการนำไปใช้บีบอัดข้อมูล ที่เน้นประสิทธิภาพการบีบอัดข้อมูล

TE 158868

Technological advances motivates the creation of a large amount of electronic documents. Most documents available world wide are written in English. Nevertheless, the number of documents in other languages is increasing quickly as well. Compression methodologies working well with English documents may not be appropriate for those in other languages. This research proposes as efficient data compression technique for Thai language which modifies the Prediction by Partial Matching method by taking into account the language characteristics. The experimental experiments show advantages over variations of Prediction by Partial Matching proposed previously, Huffman coding, and Entropy.