การปรับเปลี่ยนการอัดข้อความภาษาไทย

นางสาวเรวดี ลิมปิโชติกุล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
ภาควิชาวิศวกรรมคอมพิวเตอร์
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
พ.ศ. 2534
ISBN 974-578-965-8

IMPROVEMENT OF THAI TEXT COMPRESSION

Miss Rawedee Limpichotikul

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science

Department of Computer Engineering

Graduate School

Chulalongkorn University

1991

ISBN 974-578-965-8

หัวข้อวิทยานิพนธ์	การปรับเปลี่ยนการอัดข้อความภาษาไทย
โดย	นางสาวเรวดี ลิมปิโชติกุล
ภาควิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ จารุมาตร ปิ่นทอง
	ผู้ช่วยศาสตราจารย์ ดร. วีระ ริ้วพิทักษ์

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ ดร. ถาวร วัชราภัย)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ ไกรวิชิต ตันติเมธ)

..... อาจารย์ที่ปรึกษา
(อาจารย์ จารุมาตร ปิ่นทอง)

..... อาจารย์ที่ปรึกษาร่วม
(ผู้ช่วยศาสตราจารย์ ดร. วีระ ริ้วพิทักษ์)

..... กรรมการ
(รองศาสตราจารย์ เดือน สินธุพันธ์ประทุม)

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสีเขียวนี้เพียงแผ่นเดียว

เรวดี ลิมปิโชติกุล : การปรับเปลี่ยนการอัดข้อความภาษาไทย (IMPROVEMENT OF THAI TEXT COMPRESSION) อ.ที่ปรึกษา : อ.จารุฆาตร ปิ่นทอง อ.ที่ปรึกษาร่วม : ผศ.ดร. วีระ ริ้วพิทักษ์, 82 หน้า. ISBN 974-578-965-8

การวิจัยครั้งนี้มีจุดมุ่งหมาย เพื่อศึกษาหาวิธีการอัดข้อมูลที่มีประสิทธิภาพ และปรับเปลี่ยนวิธีการจากเดิมให้มีประสิทธิภาพเพิ่มขึ้น สำหรับนำไปใช้ในการอัดข้อความภาษาไทย จากการศึกษาการอัดข้อมูลวิธีต่าง ๆ ในปัจจุบัน พบว่าวิธีการที่มีประสิทธิภาพ และเหมาะสมกับการอัดข้อความภาษาไทย ได้แก่ วิธีการของอัลกอริทึมฮัฟแมน (Huffman Algorithm) และวิธีการของอัลกอริทึมแอลแซดดับบลิว (LZW Algorithm) การปรับเปลี่ยนวิธีการจะนำข้อมูลคำไทยที่พบบ่อยในชีวิตประจำวัน จากงานวิจัยของมหาวิทยาลัยเกษตรศาสตร์ เรื่อง การวิเคราะห์คำไทย (Thai Word Analysis) สร้างเป็นแฟ้มข้อมูลคำไทย เพื่อนำมาใช้ในวิธีการทั้งสอง

ในการทดสอบเพื่อดูประสิทธิภาพการอัดข้อมูลของวิธีการที่ปรับเปลี่ยนกับข้อมูลขนาดต่าง ๆ พบว่าวิธีการที่ปรับเปลี่ยนสามารถให้ประสิทธิภาพการอัดข้อมูลสูงขึ้นกว่าเดิม สามารถลดขนาดข้อมูลลงได้โดยเฉลี่ย 40-55 เปอร์เซนต์ ซึ่งขึ้นอยู่กับขนาดของข้อมูลและลักษณะของข้อมูลที่สอดคล้องกับแฟ้มข้อมูลคำไทย เมื่อพิจารณาเปรียบเทียบระหว่างวิธีการที่ปรับเปลี่ยนทั้งสองวิธีพบว่า วิธีการของอัลกอริทึมฮัฟแมนที่ปรับเปลี่ยนเป็นวิธีการที่มีอัลกอริทึมที่ซับซ้อนกว่า และใช้เวลาในการอัดข้อมูลมากกว่าวิธีการของอัลกอริทึมแอลแซดดับบลิวที่ปรับเปลี่ยน ดังนั้นวิธีการของอัลกอริทึมแอลแซดดับบลิวที่ปรับเปลี่ยน จึงเป็นวิธีการที่เหมาะสมสำหรับนำไปใช้ในการอัดข้อความภาษาไทย

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต เรวดี ลิมปิโชติกุล

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ลายมือชื่ออาจารย์ที่ปรึกษา

ปีการศึกษา 2533 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

RAWEDEE LIMPICHOTIKUL : IMPROVEMENT OF THAI TEXT COMPRESSION. THESIS ADVISOR : MR. JARUMATR PINTHONG. ASST.PROF. WEERA RIEWPITUK, Ph.D. 82 pp. ISBN 974-578-965-8

An objective of this research is to search for the efficient methods of data compression and improve the methods to increase the efficiency when compressing Thai text. The two efficient methods, Huffman algorithm and LZW algorithm have been found to gave a high compression efficiency and suitable for use with Thai text. An further improvement, words that are frequently found in Thai text which is obtained from Thai Word Analysis, a research of Kasetsart University, will be used to create a Thai word file for use with these two methods.

From the experiment, examples of various file sizes are presented to demonstrate the increasing performance of improved methods, give a compression efficiency of about 40-55 percents. The increasing performance depending on file sizes and Thai semantics in files. When comparing the improved methods in terms of compression time and algorithm complexity, it show that improved Huffman algorithm take more time than improved LZW algorithm because it is more complex. Therefore the improved LZW algorithm is suitable method in compressing Thai text.

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต รวี ลิมปิโชติกุล

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ลายมือชื่ออาจารย์ที่ปรึกษา

ปีการศึกษา 2533 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของ อาจารย์จารุมาตร ปิ่นทอง อาจารย์ที่ปรึกษาวิทยานิพนธ์และ ผศ.ดร. วีระ ริ้วพิทักษ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม จึงขอขอบพระคุณอาจารย์ทั้งสองมา ณ ที่นี้

ขอขอบพระคุณ รศ. ยืน ภู่วรวรรณ ที่ได้อนุเคราะห์ข้อมูลสำหรับการทำวิทยานิพนธ์ ขอขอบพระคุณสำนักคอมพิวเตอร์ มหาวิทยาลัยมหิดล ที่ได้เอื้อเฟื้ออุปกรณ์บริภัณฑ์ ขอขอบพระคุณบัณฑิตวิทยาลัย ที่ให้ทุนอุดหนุนการวิจัย ขอขอบคุณเพื่อน ๆ ที่ให้ความช่วยเหลือเป็นอย่างดี

ท้ายนี้ ผู้วิจัยใคร่ขอกราบขอบพระคุณบิดามารดาที่ได้ให้กำลังใจแก่ผู้วิจัยเสมอมา

เรวดี ลิมปิโชติกุล

สารบัญ

สารบัญตาราง

หน้า

สารบัญรูป

หน้า