

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสี่เหลี่ยมนี้เพียงแผ่นเดียว

เปรมิน จินดาวิมลเลิศ : ระบบการค้นคืนข้อความภาษาไทยโดยใช้ต้นไม้แพต (A THAI TEXT RETRIEVAL SYSTEM USING THE PAT TREE) อ.ที่ปรึกษา : ผศ. ดร. สมชาย ประสิทธิ์จตุตระกูล, 78 หน้า. ISBN 974-636-136-8.

วิทยานิพนธ์นี้นำเสนอการพัฒนาการพัฒนาระบบค้นคืนข้อความภาษาไทยโดยใช้ต้นไม้แพต เนื่องจากต้นไม้แพตจัดเก็บบรรณของสายอักขระแบบกึ่งอนันต์ ที่เรียกว่าซิสตริง ซึ่งคือลำดับย่อยของตัวอักษรต่อเนื่องกันในข้อความ จึงขจัดปัญหาการแบ่งคำในข้อความภาษาไทยที่มักกระทำไม่ได้ถูกต้องสมบูรณ์ ขั้นตอนการสร้างเริ่มแบ่งเอกสารฉบับใหม่ออกเป็นเอกสารฉบับย่อยๆ ที่มีขนาดแปรตามขนาดของหน่วยความจำหลักที่มีเหลืออยู่ในระบบ จากนั้นตัดซิสตริงที่มีจุดเริ่มต้นที่ไม่ถูกต้องกับหลักภาษาไทยเบื้องต้นออกจากขั้นตอนการจัดเก็บแล้วสร้างแพตอะเรย์ที่เก็บซิสตริงที่เหลือของแต่ละเอกสารย่อย โดยใช้ต้นไม้แพตเป็นโครงสร้างข้อมูลภายในเมื่อได้แพตอะเรย์ของทุกๆ เอกสารย่อยแล้ว จึงนำแพตอะเรย์เหล่านี้มาผสานเข้าด้วยกันกับแพตอะเรย์เดิมของเอกสารก่อนๆ เป็นแพตอะเรย์ใหม่ที่เก็บบรรณของเอกสารฉบับใหม่นั้นด้วย จากนั้นสร้างบรรณระดับที่สอง (ที่มีขนาดเพียงพอต่อการจัดเก็บในหน่วยความจำหลัก) สำหรับอ้างอิงข้อมูลในแพตอะเรย์ เพื่อเป็นการเพิ่มประสิทธิภาพการเข้าถึงข้อมูล

ขั้นตอนการสร้างที่ได้กล่าวถึงนี้ใช้เวลาการทำงานเป็น $O(k(N+n))$ โดยที่ k คือจำนวนเอกสารย่อย n คือขนาดของเอกสารฉบับใหม่ที่จะถูกเพิ่ม และ N คือขนาดของแพตอะเรย์ก่อนการเพิ่มเอกสารฉบับใหม่ นอกจากนี้ยังได้แสดงให้เห็นว่าแพตอะเรย์จะมีขนาดที่แปรผันตามขนาดของซิสตริง จากการทดลองกับข้อความภาษาไทยพบว่า อัตราการเพิ่มขนาดของแพตอะเรย์น้อยกว่า 1% เมื่อซิสตริงมีความยาวตั้งแต่ 15 ตัวอักษรเป็นต้นไป

ภาควิชา วิศวกรรม คอมพิวเตอร์
สาขาวิชา วิทยาการคอมพิวเตอร์
ปีการศึกษา 2539

ลายมือชื่อนิสิต เปรมิน จินดาวิมลเลิศ
ลายมือชื่ออาจารย์ที่ปรึกษา ประสิทธิ์จตุตระกูล
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม