

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสี่เหลี่ยมนี้เพียงแผ่นเดียว

C718342 : MAJOR COMPUTER SCIENCE
KEY WORD: PAT TREE / PAT ARRAY / THAI TEXT RETRIEVAL / TEXT RETRIEVAL

PARAMIN JINDAVIMONLERT : A THAI TEXT RETRIEVAL SYSTEM USING THE PAT TREE. THESIS ADVISOR : ASSIST. PROF. SOMCHAI PRASITJUTRAKUL, Ph.D.
78 pp. ISBN 974-636-136-8.

This thesis presents a development of Thai text retrieval system using PAT trees. By organizing semi-infinite strings (sistring) as indices in PAT trees which are subsequences of characters, it eliminates the need to perform Thai text segmentation which is usually not 100% correct. The building process begins by first dividing a new document into a set of small subdocuments whose sizes depend on available memory in the system. Next any sistring with ineligible Thai starting characters are eliminated for further consideration. Then a PAT array is built from the set of eligible sistrings by using a PAT trees as one of internal data structures. After obtaining PAT arrays for all the subdocuments, these PAT arrays are then merged with the original PAT array to form a new PAT array having indices for the new document. Finally, a second level index (whose size is sufficiently small for keeping in the available memory) is built for the entire PAT array in order to improve access time.

The whole building process takes time in $O(k(N+n))$ where k is the number of subdocuments, n is the size of the new document to be added, and N is the size of the PAT array before adding the new document. It is also shown that the size of the PAT array increases as sistrings get longer. Experimental results showed that the growth rate of PAT array's size is less than 1% when sistring is of length starting from 15 characters.

ภาควิชา..... วิศวกรรม คอมพิวเตอร์
สาขาวิชา..... วิทยาการคอมพิวเตอร์
ปีการศึกษา..... 2539

ลายมือชื่อนิสิต..... ประมวน ดินตาคิมล เลิศ
ลายมือชื่ออาจารย์ที่ปรึกษา..... ธีระชัย ประสงค์สุภาวดี
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....