

A HYBRID PATTERN MATCHING ALGORITHM

KANUMPORN ASAWASIROJ 5336468 EGTI/M

M.Sc. (TECHNOLOGY OF INFORMATION SYSTEM MANAGEMENT)

THESIS ADVISORY COMMITTEE: NOPPADOL WANICHWORANANT, Ph.D.
RANGSIPAN MARUKATAT, Ph.D. SUPAPORN KIATTISIN, Ph.D.

ABSTRACT

The pattern matching algorithms are widely used in computer science, including information retrieval, text editors, internet search engines, and biological applications. The research proposed a pattern matching algorithm by the combination of three patterns matching algorithms including Boyer-Moore-Horspool, Quick Search, and Raita algorithm, in order to produce the hybrid pattern matching algorithms, given as: Hybrid Max Shift and Reverse Hybrid Max Shift. The proposed algorithms were improved from Boyer-Moore-Horspool and Quick search algorithm by choosing the maximum of shifting value among them and applying with Raita algorithm's order comparing technique. Four datasets were used to test the proposed algorithms, given as English text, genome sequence, protein sequence, and random texts. The best- and worst- case time complexities were also presented in this research. By the experiments, the proposed algorithms were compared to the other existing algorithms. It was shown that the proposed algorithms outperform other existing pattern matching algorithms in terms of shorter average running time.

KEY WORDS: PATTERN MATCHING ALGORITHM / BOYER-MOORE-
HORSPPOOL / QUICK SEARCH / RAITA / HYBRID

103 pages

การพัฒนาอัลกอริทึมการค้นหาสายอักขระ

A HYBRID PATTERN MATCHING ALGORITHM

คนัมพร อัสวศิริโรจน์ 5336468 EGTI/M

วท.ม. (เทคโนโลยีการจัดการระบบสารสนเทศ)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์: นกมล วณิชวรนนท์, Ph.D., รังสิพรรณ มฤคทัต, Ph.D., สุภาภรณ์ เกียรติสิน, Ph.D.

บทคัดย่อ

อัลกอริทึมเปรียบเทียบรูปแบบถือว่ามีความสำคัญและแพร่หลายในวงการวิทยาศาสตร์คอมพิวเตอร์ ซึ่งอัลกอริทึมนี้สามารถพบได้ในโปรแกรมประยุกต์ต่างๆ เช่น การค้นหาสารสนเทศ, โปรแกรม Text Editor, โปรแกรมค้นหาบนอินเทอร์เน็ตและโปรแกรมทางชีววิทยา งานวิจัยนี้ได้นำเสนอวิธีการพัฒนาอัลกอริทึมเปรียบเทียบรูปแบบ โดยการใช้การผสมผสานข้อดีของอัลกอริทึม 3 อัลกอริทึมเข้าด้วยกัน ได้แก่ Boyer-Moore-Horspool, Quick Search และ Raita อัลกอริทึมที่พัฒนาขึ้นมาให้ชื่อว่า Hybrid Max Shift และ Reverse Hybrid Max Shift ทั้ง 2 อัลกอริทึมนี้จะใช้การเลือกค่าสูงสุดของการเลื่อนจากการคำนวณที่ได้มาจากอัลกอริทึม Boyer-Moore-Horspool และ Quick Search ร่วมกับการใช้ลำดับการเปรียบเทียบของอัลกอริทึม Raita การทดสอบผลจะทดสอบบนกลุ่มข้อมูล 4 ประเภท ได้แก่ ข้อมูลภาษาอังกฤษ, ข้อมูลของสายพันธุกรรม, ข้อมูลของสายโปรตีนและข้อมูลแบบสุ่ม การวิเคราะห์ประสิทธิภาพทั้งกรณีที่ดีที่สุดและแย่ที่สุดก็ได้ถูกนำมาแสดงในงานวิจัยนี้ นอกจากนี้ยังถูกนำมาเปรียบเทียบผลการทำงานร่วมกับอัลกอริทึมอื่นๆ ที่เกี่ยวข้องในงานวิจัย ผลที่ได้จากการทดลอง แสดงผลออกมาว่าอัลกอริทึมที่พัฒนามีประสิทธิภาพในการทำงานที่สูงกว่าอัลกอริทึมอื่นๆ ในแง่ของค่าเฉลี่ยเวลาที่ใช้ในการทำงาน