

**AUGMENTING PREFIXSPAN ALGORITHM TO SUPPORT  
PUBLIC PHONE MALFUNCTION PREDICTION**



**VARALUK DILOKTRAKARNKIJ**

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR  
THE DEGREE OF MASTER OF SCIENCE  
(COMPUTER SCIENCE)  
FACULTY OF GRADUATE STUDIES  
MAHIDOL UNIVERSITY  
2008**

**COPYRIGHT OF MAHIDOL UNIVERSITY**

การแต่งเติมอัลกอริทึมพรีฟิกซ์สเปนเพื่อสนับสนุนการทำนายเหตุเสียเครื่องโทรศัพท์สาธารณะ  
(AUGMENTING PREFIXSPAN ALGORITHM TO SUPPORT PUBLIC PHONE  
MALFUNCTION PREDICTION)

วราลักษณ์ คิลกตระกูลกิจ 4737223 SCCS/M

วท.ม. (วิทยาการคอมพิวเตอร์)

คณะกรรมการควบคุมวิทยานิพนธ์ : ชาญยศ ปลื้มปิติวิริยะเวช, Ph.D., ทรงศรี ตั้งศรีไพโรจน์,  
Ph.D.

บทคัดย่อ

พรีฟิกซ์สเปนอัลกอริทึมเป็นอัลกอริทึมหนึ่งของเทคนิคการขุดค้นรูปแบบเชิงลำดับ ที่ใช้สำหรับการวิเคราะห์ข้อมูลประเภทเชิงลำดับเหตุการณ์ เพื่อค้นหาความสัมพันธ์ของเหตุการณ์ที่เกิดขึ้นในชุดข้อมูลนั้น แนวความคิดหลักของพรีฟิกซ์สเปนคือการใช้เฉพาะส่วนเติมหน้าลำดับเหตุการณ์เป็นส่วนหลักในกระบวนการทำงาน พร้อมค่าความถี่ขึ้นต่ำสำหรับคัดเลือกรูปแบบผลลัพธ์ พรีฟิกซ์สเปนเป็นอัลกอริทึมที่มีประสิทธิภาพการทำงานทั้งในด้านการจัดสรรหน่วยความจำ และการตอบสนองของเวลา อย่างไรก็ตาม เราพบว่าการทำงานของพรีฟิกซ์สเปนได้ค้นหารูปแบบผลลัพธ์ที่มีความน่าสนใจได้ออกมาจำนวนมาก

วิทยานิพนธ์นี้อธิบายถึงกระบวนการทำงานที่สำคัญในการแต่งเติมพรีฟิกซ์สเปนอัลกอริทึม โดยได้นำค่าความแม่นยำขึ้นต่ำมาร่วมเป็นเงื่อนไขสำหรับการคัดเลือกรูปแบบผลลัพธ์ที่มีความน่าสนใจต่ำ และมีประโยชน์ต่อการใช้งานน้อยออกไปอย่างอัตโนมัติ นอกจากนี้กระบวนการทำงานนี้ได้รวมถึงการแต่งเติมขั้นตอนการจัดการข้อมูลก่อนและหลังเข้าสู่กระบวนการหลัก เช่น การทำความสะอาด การตรวจสอบ การจัดรูปแบบข้อมูลใหม่เพื่อรองรับพรีฟิกซ์สเปน และการคำนวณระยะเวลาการเกิดเหตุการณ์ที่สัมพันธ์กันอีกด้วย โดยกระบวนการทำงานทั้งหมดนี้ถูกนำไปใช้วิเคราะห์ชุดข้อมูลเหตุเสียเครื่องโทรศัพท์สาธารณะ

จากผลการทดสอบแสดงให้เห็นถึงการแต่งเติมพรีฟิกซ์สเปนอัลกอริทึม สามารถลดจำนวนรูปแบบผลลัพธ์ที่ไม่น่าสนใจลงและเพิ่มขีดความสามารถในการวิเคราะห์และค้นหารูปแบบความสัมพันธ์ได้ดียิ่งขึ้น งานวิจัยนี้นำมาซึ่งประโยชน์ต่อการวางแผนงานซ่อมเหตุเสียเครื่องโทรศัพท์สาธารณะ โดยมีเป้าหมายในการลดจำนวนของการเกิดเหตุเสียที่มีความสัมพันธ์กัน รวมถึงลดค่าใช้จ่ายในการเดินทางเวลา และอัตราค่าล้างสำหรับการเข้าตรวจซ่อมเหตุเสียเครื่องโทรศัพท์สาธารณะ

AUGMENTING PREFIXSPAN ALGORITHM TO SUPPORT PUBLIC PHONE MALFUNCTION PREDICTION

VARALUK DILOKTRAKARNKIJ 4737223 SCCS/M

M.Sc. (COMPUTER SCIENCE)

THESIS ADVISORS: CHARNYOTE PLUEMPITWIRIYAWAJ, Ph.D., SONGSRI TANGSRIPAHOJ, Ph.D.

ABSTRACT

PrefixSpan is a sequential pattern mining algorithm used to analyze a set of event transactions and discover relationships among them. Its main idea is to examine only the prefix sub-sequences with minimum support. PrefixSpan is efficient in terms of memory requirement and response time. However, it comes up with a large number of sequential patterns, only a few of which are interesting.

This thesis describes the underlying mechanism which augments the PrefixSpan algorithm. Specifically, a new mechanism which additionally considers minimum confidence of patterns to automatically filter out uninteresting and useless discovered patterns was developed. In addition, the mechanism includes data pre-processing and data post-processing such as data cleaning, verification, reformatting, and time duration calculation. The mechanism was evaluated on a public phone malfunction dataset.

The experimental results showed that the augmentation of PrefixSpan can reduce the number of uninteresting patterns and increase the discovery capacity. This research contributes to public phone maintenance planning whose goal is to minimize the number of malfunctions, the transportation expenses, the time and the manpower for maintenance.

KEY WORDS: PREFIXSPAN / SEQUENTIAL PATTERN MINING / PUBLIC PHONE MALFUNCTION PREDICTION

64 pp.