

ชื่อเรื่องวิทยานิพนธ์	การปรับปรุงสมรรถนะ โดยการลดมิติในการตรวจหารหัสคำสั่งซ้ำแบบไอโซเอสที
ผู้เขียน	นายณรงค์ฤทธิ์ สุวรรณรักษ์
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต (วิศวกรรมคอมพิวเตอร์)
อาจารย์ที่ปรึกษาวิทยานิพนธ์	อาจารย์ ดร. พฤษภ บุญญา

บทคัดย่อ

ในปัจจุบันได้มีการพัฒนาซอฟต์แวร์ประเภทต่างๆ ที่ใช้สำหรับการระบุรหัสคำสั่งซ้ำภายในโปรแกรม โดยที่มีจุดประสงค์หลักคือทำให้ผู้พัฒนาสามารถตัดหรือลดจำนวนรหัสคำสั่งบางส่วนที่ไม่จำเป็นออกไปได้ หรือ ช่วยให้ผู้พัฒนามองเห็นปัญหา ข้อบกพร่องจากจำนวนรหัสคำสั่งที่ซ้ำกัน และ นำมาซึ่งข้อมูลเชิงลึกเพื่อที่จะนำมาใช้ในการวิเคราะห์ ออกแบบ หรือพัฒนาโปรแกรมต่อไป

ซึ่งได้มีงานวิจัยที่นำเอาหลักการหาความสัมพันธ์จากโครงสร้างวากยสัมพันธ์ หรือ การวิเคราะห์จากความหมายของสัญลักษณ์ในภาษา มาช่วยในการตรวจสอบหารหัสคำสั่งซ้ำ โดยในส่วนของรูปแบบการวิเคราะห์จากโครงสร้างวากยสัมพันธ์นั้น ได้มีการนำเอาโครงสร้างข้อมูลแบบต้นไม้ที่แสดงถึงโครงสร้างวากยสัมพันธ์ภายในโปรแกรมคือ แอ็บสแตรคซินแทกซ์ทรี หรือ เอเอสที มาใช้ตรวจสอบรหัสคำสั่งซ้ำ รวมไปถึงพัฒนาอัลกอริทึมโดยใช้เทคนิคการปรับโหนดจากโครงสร้างต้นไม้ ให้อยู่ในรูปของแคแรกเทอร์สตริงเวกเตอร์ เพื่อลดปัญหาการระบุความคล้ายกันของโครงสร้างต้นไม้ที่ซับซ้อน เป็นการหาความคล้ายกันของเวกเตอร์แทน

อย่างไรก็ตามการวิเคราะห์ในลักษณะนี้ จำนวนประเภทโหนดจากโครงสร้างต้นไม้ที่ได้ จะมีผลต่อความถูกต้องในการระบุความคล้ายจากเวกเตอร์ และจำนวนโหนดที่มากเกินไปก็ส่งผล ต่อขนาดมิติของเวกเตอร์ รวมไปถึงระยะเวลาในการประมวลผลเช่นกัน

ดังนั้นในงานวิจัยนี้จึงนำเสนอเกี่ยวกับการลดขนาดเอเอสที โดยการลดจำนวนโหนดเป็นหลักเพื่อลดจำนวนมิติของแคแรกเทอริสติกเวกเตอร์ โดยที่ยังคำนึงถึงความถูกต้องของโครงสร้าง วากยสัมพันธ์และลักษณะการทำงานของโปรแกรม ซึ่งจากการทดลองพบว่าประสิทธิภาพจาก ตรวจสอบรหัสคำสั่งซ้ำของเอเอสทีที่ลดขนาดอยู่ในระดับที่น่าพอใจ เมื่อวัดผลจากความถูกต้องของ กลุ่มรหัสคำสั่งซ้ำที่ได้ใกล้เคียงกับวิธีเดิมแต่ใช้ระยะเวลาลดลง

Thesis Title	Performance Improvement by Dimension Reduction in AST-Based Code-Duplication Detection
Author	Mr. Narongrit Suwannarak
Degree	Master of Engineering (Computer Engineering)
Thesis Advisor	Dr. Pruet Boonma

ABSTRACT

In general, several techniques have been developed for code duplication, such as clone code detection in programs. Its main purpose is used to identify redundant code to gain insight into program design or code modifications. Some approaches for clone detection present algorithm identifying similar subtree to representations of similar source code and using numerical vectors to approximate tree structures, for reduce the tree similarity problem to detecting similar vectors. Where the dimension of vectors depend on count occurrences of a specific tree pattern, i.e. the number of node kinds in a parse tree. So we present a performance improvement for code duplication detection by using tree reduction technique that affect to dimensional reduction. The proposed approach was compared with the one without dimensional reduction and shows that both produce similar set of duplicated code while the proposed approach can perform faster.