

งานวิจัยนี้นำเสนออัลกอริทึมใหม่สำหรับการสืบค้นเซตของรูปแบบสตริงย่อยแบบความยาวสูงสุดที่เกิดบ่อยที่เรียกว่า FM จากสตริงนำเข้าใดๆ โดยที่ค่าความถี่ขั้นต่ำที่กำหนดให้ FM คือ เซตของรูปแบบสตริงย่อยที่มีขนาดความยาวสูงสุดที่มีค่าความถี่ผ่านค่าความถี่ขั้นต่ำที่กำหนดให้นี้ งานวิจัยนี้ใช้โครงสร้างข้อมูลใหม่ที่เรียกว่าซัพฟิกซ์ทรีแบบมีความถี่ หรือ FST ซึ่งรับประกันความครบถ้วนและถูกต้องในการเจนนับรูปแบบสตริงย่อย อัลกอริทึมใหม่ทำการสืบค้น FM แบบมีประสิทธิภาพโดยใช้คุณสมบัติบางประการของโครงสร้าง FST

ABSTRACT

187550

This research proposes a novel algorithm for mining a set of frequent max substring patterns called FM from an input string. At a given threshold, FM is a set of maximum-length substring patterns with their frequencies at least at the threshold value. The research uses the new data structure called the frequent suffix trie or FST to assure exhaustive enumeration of substring patterns. The new algorithm efficiently searches for FM using some FST properties.