

ในปัจจุบันที่อยู่บนซองจดหมายมีทั้งที่อยู่ที่สมบูรณ์ และที่อยู่ที่ไม่สมบูรณ์ ที่อยู่ที่ไม่สมบูรณ์เกิดจากการสะกดผิด ส่วนประกอบที่อยู่ขาดหาย และส่วนประกอบของที่อยู่อาจมีชื่อซ้ำกัน ซึ่งทำให้เกิดการค้นหารหัส ผู้รับมีความผิดพลาด วิทยานิพนธ์นี้เสนอการแปลความหมายของที่อยู่บนซองจดหมายไทยเพื่อค้นหารหัสผู้รับ โดยใช้การค้นหารหัสผู้รับ 2 วิธี วิธีการแรกคือการเปิดตารางที่อยู่ในกรณีที่อยู่สมบูรณ์ และวิธีการที่ 2 คือ การค้นหาส่วนประกอบที่อยู่ในกรณีของที่อยู่ที่ไม่สมบูรณ์ โดยวิธีที่สองทำการค้นหาส่วนประกอบที่อยู่ทีละระดับ โดยเริ่มจาก ประเทศไทย จังหวัด อำเภอ และตำบลตามลำดับโดยพิจารณาว่าชื่อในแต่ละระดับคือ โหนด ซึ่งวิธีการที่ 2 แบ่งออกเป็น 3 ส่วน ส่วนแรกคือการใช้กรีดีอัลกอริทึมร่วมกับการเข้าถึงข้อมูลโดยตรง เพื่อใช้ในการค้นหาส่วนประกอบที่อยู่ในกรณีของส่วนประกอบที่อยู่สมบูรณ์ โดยใช้ค่าความเหมือนสะสมของสายอักขระในการเลือกโหนดต่อไป ซึ่งเลือกจากความเหมือนสะสมที่มีค่ามากที่สุด ถ้าค่าความเหมือนสะสมเท่ากันทำการค้นหาในโหนดลูกเหล่านั้นทั้งหมด ส่วนที่สองคือการใช้กรีดีอัลกอริทึมร่วมกับการเทียบสายอักขระ เพื่อใช้ในการค้นหาส่วนประกอบที่อยู่ที่ไม่สมบูรณ์ โดยทำการเทียบสายอักขระโดยประมาณจากตำแหน่งขวาสุดของข้อความที่อยู่ผู้รับมายังซ้ายสุดเพื่อค้นหาสายอักขระ ส่วนสุดท้ายทำการตัดสินใจเลือกทิศทางในเส้นทางที่มีค่าความเหมือนสะสมมากที่สุด แต่ถ้าค่าความเหมือนสะสมสูงสุดมีมากกว่าหนึ่งเส้นทางจะถือว่าไม่สามารถแปลความหมายที่อยู่ได้ ผลการทดสอบการแปลความหมายของที่อยู่บนซองจดหมายภาษาไทย ซึ่งใช้กลุ่มตัวอย่าง 900 ซอง ระบบสามารถค้นหารหัสผู้รับได้ถูกต้องร้อยละ 91

Nowadays, addresses on the envelopes can be both perfect and imperfect because of misspelling, components missing, and component repeating. Therefore, final delivery point codes often have errors. This thesis proposes an automatic interpretation of Thai mail addresses, which is used to search for final delivery point codes by using two methods: (1) using look up table in case of perfect address, and (2) searching for address components in case of imperfect address. The searching method utilizes the tree data structure by considering address database as a tree and searching the tree in each level started from the highest level; that is the search is performed in the country level, province level, district level and sub-district level, respectively. There are 3 parts in the second method. The first part uses greedy algorithm together with direct data access to search for perfect address component in the level. Then, the nodes with the highest value of accumulative similarity are selected. The second part uses greedy algorithm together with string matching to search for imperfect address components by approximate string matching, which starts searching from right to left of the string to find the character pattern. In the final part, the path that has highest value of accumulative similarity is selected as the final delivery point. However, if there are more than one path with maximum score, all paths are rejected and the address is considered to have no final delivery point. The experimental results on 900 addresses samples show 91 percent accuracy.