

ชื่อเรื่องวิทยานิพนธ์

การรู้จำตัวพิมพ์อักษรธรรมล้านนา โดยวิธีเค-เนียร์
เรสเนเบอร์และคอนดิชันนัลแรนคอมฟิลด์

ผู้เขียน

นางสาวชุตินา เชื้อพูล

ปริญญา

วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)

อาจารย์ที่ปรึกษาวิทยานิพนธ์

รองศาสตราจารย์ ดร. จิรยุทธ ไชยจารุณิช

บทคัดย่อ

วิทยานิพนธ์นี้นำเสนอการรู้จำตัวพิมพ์อักษรธรรมล้านนาโดยวิธีเค-เนียร์เรสเนเบอร์และคอนดิชันนัลแรนคอมฟิลด์ เพื่อแปลงภาพเอกสารล้านนาให้เป็นตัวอักษรที่สามารถแก้ไขได้บนเครื่องคอมพิวเตอร์ ซึ่งช่วยรักษาความรู้ในเอกสารโบราณให้คงอยู่ต่อไป และการค้นหาเอกสารทำได้ง่ายขึ้น โดยเริ่มจากการปรับปรุงคุณภาพเอกสารให้เหมาะสม ด้วยการกำจัดสัญญาณรบกวนและการปรับความเอียงของเอกสาร จากนั้นตัดแบ่งเอกสารให้เป็นบล็อกกลุ่มตัวอักษรเพื่อรู้จำตัวอักษรทีละบล็อก เนื่องจากการตัดแบ่งให้เป็นตัวอักษรอาจทำให้ตัวหนังสือที่ลายเส้นขาดหายจะถูกแยกเป็นหลายส่วน ซึ่งส่งผลให้การรู้จำไม่ถูกต้อง โดยในการรู้จำนี้ใช้วิธีเค-เนียร์เรสเนเบอร์ในการจำแนกประเภทตัวอักษร อย่างไรก็ตามวิธีการรู้จำนี้ยังมีโอกาสผิดพลาด เนื่องจากชุดข้อมูลฝึกสอนไม่ครอบคลุมประเภทของบล็อกตัวอักษรทั้งหมด จึงใช้วิธีเทมเพลตแมตชิ่งรู้จำตัวอักษรซ้ำอีกครั้ง และใช้วิธีคอนดิชันนัลแรนคอมฟิลด์ เพื่อตรวจสอบความถูกต้องของผลลัพธ์การรู้จำ โดยอาศัยลักษณะการปรากฏร่วมกันของตัวอักษร ซึ่งมีการกำหนดคุณสมบัติต่างๆ เพื่อช่วยในการพิจารณา โดยไม่จำเป็นต้องมีผู้เชี่ยวชาญในการวิเคราะห์กฎของภาษา ผลทดสอบการรู้จำมีอัตราความถูกต้องในการรู้จำร้อยละ 93.93

Thesis Title	Lanna Dharma Printed Character Recognition Using K-Nearest Neighbor and Conditional Random Fields Methods
Author	Ms. Chutima Chueaphun
Degree	Master of Science (Computer Science)
Thesis Advisor	Associate Professor Dr. Jeerayut Chaijaruwanich

ABSTRACT

This thesis proposes Lanna Dharma printed character recognition using k -Nearest Neighbor and Conditional Random Fields methods. The thesis aims to convert the Lanna image into text to preserve the ancient Lanna knowledge and to ease accessing document. Image preprocessing is first performed to enhance, denoise and adjust skew of images. After that, images are segmented into blocks which may contain several closely consecutive characters. Next, k -Nearest Neighbor algorithm is used to firstly classify the block images. Then, the uncertain classification with large dissimilarity is fine-tuned by template matching technique. Finally, a sequence of characters which is the output of k -Nearest Neighbor and template matching is contextually reclassified again by Conditional Random Fields. The experimental results show our best recognition accuracy rate at 93.93%.