193136

วิทยานิพนธ์ฉบับนี้นำเสนอการค้นคืนข้ามภาษา สำหรับคำทับศัพท์ภาษาไทย/อังกฤษโดยได้ใช้วิธีการของนิวรอลเน็ตเวิร์กในการเข้ารหัสคำ และใช้ขั้นตอนการตัดพยางค์ของรหัสเสียง วิธีการที่นำเสนอช่วยให้สามารถค้นคืนคำทับศัพท์ข้ามภาษาได้โดยไม่ต้องอาศัยพจนานุกรม

ในการค้นคืนข้ามภาษาโดยไม่อาศัยพจนานุกรมนั้นจำเป็นต้องใช้หลักการเข้ารหัสซึ่งเป็นสัญลักษณ์แทนเสียงอ่านของคำและประกอบด้วยรหัสเสียงของแต่ละตัวอักษรของคำมาเรียงต่อกัน ในการที่จะทราบว่าตัวอักษรที่กำลังสนใจในคำนั้นให้รหัสเสียงใดจำเป็นต้องอาศัยการพิจารณาตัวอักษรข้างเคียงด้วย ดังนั้นการเข้ารหัสคำสามารถจัดได้ว่าเป็นปัญหาการจำแนกอย่างหนึ่ง ด้วยเหตุนี้จึงได้นำวิธีการนิวรอลเน็ตเวิร์กมาใช้ในการเข้ารหัสคำ แต่เนื่องจากว่ารหัสคำของคำไทยและอังกฤษที่มีเสียงอ่านตรงกัน อาจมีความแตกต่างกันบ้าง จึงได้ใช้ขั้นตอนการเปรียบเทียบแบบประมาณสำหรับการค้นคืนคำที่มีเสียงอ่านคล้ายกันมากที่สุด จากผลการทดลองด้วยวิธี K-fold cross validation พบว่าเมื่อได้ปรับปรุงนิวรอลเน็ตเวิร์ก สามารถให้ผลการค้นคืนในแบบที่ 1 ด้วยตัววัด F1 เป็น 83.28% สำหรับกรณีคำไทยทับศัพท์คำอังกฤษและให้ผลการค้นคืน F1 90.54% สำหรับคำอังกฤษทับศัพท์คำไทยที่ค่าความแตกต่างของรหัสเสียงเป็น 0

193136

This thesis presents Thai/English cross-language transliterated word retrieval by using neural networks and syllable segmentation of phonetic codes. The proposed method enables the transliterated word retrieval without using the dictionary.

Without dictionary, the phonetic code is employed for cross-language retrieval. The phonetic code of a word represents the sound of the word and it consists of a sequence of phonetic codes of characters in the word. In order to determine the code of a particular character, it is necessary to consider its surrounding characters. Hence this problem can be identified as a classification problem. For this reason, neural networks are used in phonetic encoding. However, as the codes generated from a pair of corresponding Thai/English words are sometimes slightly different, the approximate string matching is applied to determine of character editing. The experimental results, using K-fold cross validation, show that the F1-measure values are 83.28% for Thai/English cross-language translitarated and 90.54% for English/Thai cross-language transliterated with zero distance between phonetic codes.