

T167230

นาย ศิริพจน์ สุรบถโสภณ : การค้นคืนข้ามภาษาสำหรับคำทับศัพท์ภาษาไทย/อังกฤษด้วย
วิธีการนิรวลเน็ตเวิร์ก แบบจำลองฮิดเดินมาร์คอฟ และขั้นตอนวิธีเชิงพันธุกรรม.

(THAI/ENGLISH CROSS-LANGUAGE TRANSLITERATED WORD RETRIEVAL USING
NEURAL NETWORKS, HIDDEN MARKOV MODELS, AND GENETIC ALGORITHMS)

อ. ที่ปรึกษา : ผศ.ดร.บุญเสริม กิจศิริกุล, 61 หน้า. ISBN 974-17-6383-2.

วิทยานิพนธ์ฉบับนี้นำเสนอการการค้นคืนข้ามภาษาสำหรับคำทับศัพท์ภาษาไทย/อังกฤษโดย
ใช้วิธีการนิรวลเน็ตเวิร์กและแบบจำลองฮิดเดินมาร์คอฟในการเข้ารหัสคำ และใช้ขั้นตอนวิธีเชิง
พันธุกรรม เพื่อเพิ่มความถูกต้องของการค้นคืน วิธีการที่นำเสนอช่วยให้สามารถค้นคืนคำทับศัพท์ข้าม
ภาษาได้โดยไม่ต้องอาศัยพจนานุกรม

ในการค้นคืนข้ามภาษาโดยไม่อาศัยพจนานุกรมนั้นจำเป็นต้องใช้หลักการเข้ารหัสซึ่งเป็น
สัญลักษณ์แทนเสียงอ่านของคำและประกอบด้วยรหัสเสียงของแต่ละตัวอักษรของคำมาเรียงต่อกัน ใน
การที่จะทราบว่าตัวอักษรที่กำลังสนใจในคำนั้นให้รหัสเสียงใดจำเป็นต้องอาศัยการพิจารณาตัวอักษร
ข้างเคียงด้วย ดังนั้นการเข้ารหัสคำสามารถจัดได้ว่าเป็นปัญหาการจำแนกอย่างหนึ่ง ด้วยเหตุนี้จึงได้นำ
วิธีการนิรวลเน็ตเวิร์กและแบบจำลองฮิดเดินมาร์คอฟมาใช้ในการเข้ารหัสคำ แต่เนื่องจากว่ารหัสคำ
ของคำไทยและอังกฤษที่มีเสียงอ่านตรงกัน อาจมีความแตกต่างกันบ้าง จึงได้ใช้ขั้นตอนวิธีเชิง
พันธุกรรมเพื่อหาต้นทุนการแก้ไขอักขระที่ใช้ในเทคนิคการเปรียบเทียบแบบประมาณสำหรับการค้นคืน
คำที่มีเสียงอ่านคล้ายกันมากที่สุด จากผลการทดลองด้วยวิธี K-fold cross validation พบว่าเมื่อใช้
นิรวลเน็ตเวิร์กร่วมกับขั้นตอนวิธีเชิงพันธุกรรมสามารถให้ผลการค้นคืน F1 ได้ประมาณ 90% และเมื่อ
ใช้แบบจำลองฮิดเดินมาร์คอฟกับขั้นตอนวิธีเชิงพันธุกรรมสามารถให้ผลการค้นคืน F1 ได้ประมาณ
80%

ภาควิชา วิศวกรรมคอมพิวเตอร์.....ลายมือชื่อนิสิต.....
สาขาวิชา วิศวกรรมคอมพิวเตอร์.....ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา 2547.....ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

TE167230

4570562421 : MAJOR COMPUTER ENGINEERING

KEY WORD: TRANSLITERATED WORD / INFORMATION RETRIEVAL / NEURAL NETWORKS / HIDDEN MARKOV MODELS / GENETIC ALGORITHM

SIRIPOD SURABOTSOPHON : THAI/ENGLISH CROSS-LANGUAGE TRANSLITERATED WORD RETRIEVAL USING NEURAL NETWORKS, HIDDEN MARKOV MODELS, AND GENETIC ALGORITHMS. THESIS ADVISOR : ASST. PROF. BOONSERM KIJSIRIKUL, Ph.D., 61 pp. ISBN 974-17-6383-2.

This thesis presents Thai/English cross-language transliterated word retrieval by using Neural Networks and Hidden Markov Models for encoding words and using the Genetic Algorithm for improving the efficiency of the retrieval. 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 and Hidden Markov Models are used in phonetic encoding. However, as the codes generated from a pair of corresponding Thai/English words are sometimes slightly different, the Genetic Algorithm is applied to determine the appropriate cost of character editing used in approximate string matching. The experimental results, using K-fold cross validation, show that the F1-measure of 90% can be obtained when using Neural Networks and the Genetic Algorithm, and of 80% when using Hidden Markov Models and the Genetic Algorithm.

Department Computer Engineering..... Student's.....
Field of study Computer Engineering..... Advisor's.....
Academic year 2004..... Co-advisor's.....