

การสร้างโมเดลคำหลักแบบอัตโนมัติ สำหรับระบบค้นหาคำหลัก  
จากเสียงพูดแบบหลายโดเมน

**Automatic Keyword Model Generation for  
Multi-domain Keyword Spotting**

นายสมชาย สิริวัฒนวงศ์ชัย, รัชฎา คงคะจันทร์

**บทคัดย่อ**

วิทยานิพนธ์ฉบับนี้ นำเสนอวิธีการสร้างโมเดลคำหลักแบบอัตโนมัติสำหรับการค้นหาคำหลักจากสัญญาณเสียงแบบหลายโดเมน โดยใช้โมเดลฮิดเดนมาร์คอฟ (Hidden Markov Model : HMM) ในการจำลองหน่วยเสียง (Acoustic Model) โมเดลของคำหลักที่สร้างขึ้นโดยอัตโนมัตินี้สร้างจากฐานข้อมูลที่เป็นข้อความซึ่งง่ายต่อการเพิ่มเติมและปรับปรุง เพราะเมื่อมีคำหลักใหม่ที่ต้องการค้นหาเกิดขึ้น ก็เพียงแค่เพิ่มคำหลักใหม่พร้อมด้วยลำดับของหน่วยเสียงของคำหลักนั้น (Phoneme Sequence) ลงในฐานข้อมูล โดยไม่ต้องฝึกฝนโมเดลคำหลักใหม่ในทุกครั้งที่มีการเพิ่มคำหลักใหม่เข้ามาในโดเมน ในวิทยานิพนธ์ฉบับนี้ ได้มีการประเมินประสิทธิภาพของวิธีการด้วยการใช้ข้อมูลเสียงจากโดเมนที่แตกต่างกันจำนวน 3 โดเมนคือ โดเมนของการลงทะเบียนเรียนของนักศึกษา มหาวิทยาลัยธรรมศาสตร์ โดเมนของการจองห้องพักภายในโรงแรม และโดเมน LOTUS speech corpus ที่พัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) ข้อมูลเสียงที่ใช้ในแต่ละโดเมนจะถูกแบ่งเป็นสองส่วนสำหรับฝึกฝนและประเมินผลตามลำดับ ซึ่งผลลัพธ์ที่ได้แสดงให้เห็นว่าวิธีการที่กำหนดสามารถค้นหาคำหลักจากข้อมูลเสียงได้ถูกต้องเพิ่มขึ้นเมื่อเทียบกับงานวิจัยที่ผ่านมา

## **Abstract**

This thesis proposes a new method to automatically generate keyword models for Multi-domain Keyword Spotting System (KWS). The new method generates a keyword model that is formed by combining the phoneme-based HMM (Hidden Markov Model) from text file. It has benefits in two points i.e. no need for re-training and ease in maintenance. When the user adds a new keyword to the system, the word is used as index and then be searched in the domain-dependent lexicon database for its phoneme sequences. The new keyword model is formed by combining the tied-phone of phoneme-based HMM together and then be kept for the next spotting. In this thesis, three domains of continuous speech data are employed. The first domain is about course registration of Thammasat University's students. The second one is the hotel reservation. The last one is the LOTUS speech corpus created by National Electronics and Computer Technology Center. These speech data are divided into two groups for training and spotting. The experimental results reveal that the proposed method can improve the precision and recall with significance.