

ชื่อวิทยานิพนธ์	การสกัดวลีสำนวนภาษาไทยด้วยโครงข่ายประสาทเทียม
ผู้เขียน	นางสาวกิริติ บุญโชติ
สาขาวิชา	การจัดการเทคโนโลยีสารสนเทศ
ปีการศึกษา	2548

บทคัดย่อ

งานวิจัยนี้เสนอเทคนิคการสกัดวลีสำนวนภาษาไทยโดยใช้โครงข่ายประสาทเทียม โดยวัตถุประสงค์ของการสกัดคือการค้นหาคำหรือวลีสำนวนที่สามารถนำมาเป็นตัวแทนเอกสารที่ดีได้ ด้วยในปัจจุบัน ปริมาณของสารสนเทศมีอยู่อย่างมากมายในเครือข่ายอินเทอร์เน็ต ก่อให้เกิดความต้องการระบบการค้นหาสารสนเทศที่มีความชาญฉลาด โดยในงานวิจัยนี้เป็นการนำเสนอเกี่ยวกับแนวคิดของการสกัดหรือการค้นหาใจความสำคัญของเอกสาร ด้วยการกำหนดระดับความสำคัญหรือค่าน้ำหนักของคำหรือวลีที่มีความสัมพันธ์กับไวยากรณ์และโครงสร้างทางภาษาศาสตร์ โดยสำหรับเอกสารภาษาไทยนั้นสิ่งที่ต้องทำเป็นอันดับแรกคือการตัดแยกคำหรือวลี โดยแต่ละคำหรือวลีที่ผ่านการตัดคำเชิงความหมายแบบกำกับหน้าที่คำแล้วนั้น จะเข้าสู่การให้ค่าน้ำหนักคุณสมบัติของคำหรือวลี ข้อมูลค่าน้ำหนักคำหรือวลีในการเรียนรู้และทดสอบของโครงข่ายประสาทเทียมมีด้วยกัน 5 ค่าน้ำหนัก คือ ค่าน้ำหนักความถี่การเกิดขึ้น ค่าน้ำหนักตำแหน่งในประโยค ค่าน้ำหนักตำแหน่งในย่อหน้า ค่าน้ำหนักตำแหน่งในเอกสาร และค่าน้ำหนักหน้าที่คำ โดยในงานวิจัยนี้เลือกใช้โครงข่ายประสาทเทียมแบบเพอร์เซพตรอนหลายชั้นที่มีการเรียนรู้แบบแพร่ย้อนกลับ ร่วมกับกฎและคลังข้อมูลในการเรียนรู้และจำแนกคำหรือวลีสำนวน โดยการทดสอบแนวคิดด้วยการเลือกใช้คุณสมบัติของคำหรือวลีทั้ง 5 คุณสมบัติและจำนวนโหนดในชั้นซ่อนที่แตกต่างกันออกไป เพื่อหาคุณสมบัติที่เหมาะสมต่อการค้นหาสารสนเทศ ในด้านของระยะเวลาในการประมวลผลในการเรียนรู้และทดสอบของโครงข่ายประสาทเทียมนั้น ระยะเวลาที่เร็วและช้าที่สุดในการทดสอบทั้งหมด คือ 5.7 และ 14.4 วินาที ตามลำดับ โดยเป็นการทดสอบบนเครื่องคอมพิวเตอร์ที่มีหน่วยประมวลผล Pentium Centrino และหน่วยความจำ 256 MB ซึ่งจากการทดสอบในด้านค่าความถูกต้องของการสกัดคำหรือวลีสำนวนนั้น พบว่า การเลือกใช้คุณสมบัติของคำหรือวลีที่เหมาะสมในการสกัดและการให้ค่าน้ำหนักดังแนวคิดนี้ สามารถทำให้สกัดคำหรือวลีสำนวนในเอกสารภาษาไทยได้แม่นยำถึง 90.260% สำหรับเอกสารชุดฝึกสอนและชุดทดสอบที่เป็นเอกสารเดียว 79.09% สำหรับเอกสารชุดฝึกสอนและชุดทดสอบที่เป็นเอกสารโดเมนเดียวกัน และ 82.80% สำหรับเอกสารชุดฝึกสอนและชุดทดสอบที่เป็นเอกสารคนละโดเมน โดยเทคนิคใน

งานวิจัยนี้สามารถนำมาประยุกต์ใช้กับเอกสารภาษาอังกฤษได้ เพราะเอกสารภาษาอังกฤษนั้นมีรูปแบบการเขียนที่คล้ายคลึงกับเอกสารภาษาไทย

คำสำคัญ : การสัคดีสำคัญ, เอกสารภาษาไทย, โครงข่ายประสาทเทียม

Thesis Title	Thai Keyphrases Extraction Using Artificial Neural Network
Author	Miss Keerati Boonchote
Major Program	Management of Information Technology
Academic Year	2005

ABSTRACT

This thesis presents the Thai Keyphrases Extraction technique using artificial neural network. The purpose of the extraction is to find the keywords that represent the document. Recently, the huge quantity of information in the Internet generates a demand for an intelligence information retrieval system. To respond to the demand, this thesis proposes the idea of the extraction of Thai keywords and keyphrases using Term Weighting in relation to the grammar and structure of Thai language. The Thai document is first segmented using basic Thai word segmentation with semantics. Each segmented word is then passed through the process of feature extraction which produces five features used in the subsequent neural network process. These features include word frequency, position of sentence, position of paragraph, position of document, and part of speech. A multi-layer perceptron with back-propagation supervised training artificial neural network together with rules and corpus are used to train and classify the keywords and keyphrases. Different combinations of features and document types were experimented to exercise the effectiveness of this information retrieval technique. Each combination requires different CPU time, the best case and the worst case are 57 and 144 seconds respectively on 1.5 GHz Pentium Centrino with 256 Mbytes of RAM. It is found that with the same document for training and testing, the accuracy of up to 90.26% can be achieved. For the document within the same domain, the best accuracy is 79.09%. The experiment was extended to cover the case of the document with difference domain, and the accuracy of up to 82.80% can be achieved. Moreover, it is found that this Thai keyphrases extraction technique would be applied to English document since it basically shares the same writing styles with Thai document.

Keyword : Keyphrases Extraction , Thai Document , Neural Network