**ชื่อเรื่องการค้นคว้าแบบอิสระ** การแยกวลีภาษาไทยสำหรับการประมวลผลภาษาธรรมชาติ

**ผู้เขียน** นางสาวเพชรรัช ศุภชาลาพร

**ปริญญา** วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)

**อาจารย์ที่ปรึกษาการค้นคว้าแบบอิสระ** อาจารย์ ดร.รัฐสิทธิ์ สุขะหุต

## บทคัดย่อ

การค้นคว้าแบบอิสระ เรื่องการแยกวลีภาษาไทยสำหรับการประมวลผลภาษาธรรมชาติ มีวัตถุประสงค์ เพื่อศึกษาการวิเคราะห์กลุ่มคำหรือวลีในภาษาไทย โดยใช้ฐานความรู้เกี่ยวกับหน้าที่ของคำ และฐานความรู้โครงสร้างของวลีตามหลักไวยากรณ์ภาษาไทย และสร้างโปรแกรมที่มีความสามารถในการแยกโครงสร้างของวลีในภาษาไทยตามหลักไวยากรณ์ภาษาไทย โดยระบบงานจะมีการทำงานแบ่งเป็น 3 ขั้นตอนใหญ่ๆ ได้แก่ขั้นตอนการวิเคราะห์โครงสร้างวลีในภาษาไทย และขั้นตอนการแยกโครงสร้างวลี หรือการแบ่งข้อความที่ต่อเนื่องกันออกเป็นวลีประเภทต่างๆ ตามโครงสร้างของวลี โดยในขั้นตอนนี้จะใช้คลังข้อความประโยคภาษาไทยที่กำกับหมวดของคำแล้วที่มีชื่อว่าออคิดคอร์ปัส ซึ่งเป็นผลของงานวิจัยการวิเคราะห์หน้าที่คำของศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ ส่วนขั้นตอนสุดท้ายเป็นการแสดงผลลัพธ์วลีภาษาไทยโดยมีการเรียงลำดับวลีในประโยคที่ถูกต้องตามหลักไวยากรณ์ไทยแล้ว

ระบบนี้ถูกพัฒนาภายใต้ระบบปฏิบัติการวินโดวส์เอ็กซ์พี โปรแกรมจาวาเวอร์ชัน 1.4.2_03 และโปรแกรมแปลภาษาเอเอ็นทีแอลอาร์ สำหรับสร้างกฎไวยากรณ์โครงสร้างวลีของภาษาไทย

ผลการวิจัย คือ สามารถแยกวลีภาษาไทยออกเป็นวลีประเภทต่างๆแต่ละประเภทได้โดยใช้โปรแกรมแยกวลี ทำให้มีความสะดวกเพิ่มมากขึ้น ซึ่งประสิทธิภาพของกฎการแยกวลีภาษาไทยมีความถูกต้องร้อยละ 78.70

**Reseach Title** Thai Phrase Segmentation for Natural Languages Processing

**Author** Miss. Phetcharat Suphachalaphorn

**Degree** Master of Science (Computer Science)

**Reseach Advisor** Lecturer Dr. Rattasit Sukhahuta

## ABSTRACT

This Independent Study, "Thai Phrase Segmentation for Natural Languages Processing", has two main objectives.The objectives are to analyze Thai phrase by using Thai language grammar and part of speech knowledge base and to implement a program which will be able to segment Thai phrase structure according to the Thai language grammar. There are 3 steps involved in the system.The first phase is the analysis of Thai phrase structure.The second phase is phrase segmentation or separate words from continuous text to form a phrase. In this step will be used Orchid Corpus, which is developed by NECTEC. Finally, the resulting Thai phrase ordered correctly according to Thai language grammar will be outputted.

This system runs on Microsoft Windows XP operating system with Java Program 1.4.2_03 and Thai phrase structure was created by ANTLR Program.

The research has shown that it does improve expedient ability and can separate Thai phrases into categories using phrase segmentation program. Its accuracy is 78.70%.