การตัดคำกำกวมในข้อความภาษาไทยด้วยการโปรแกรมตรรกะเชิงอุปนัย

นางสาว ชมภูนุช คุปติวุฒิ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2542

ISBN 974-332-872-6

SEGMENTATION OF AMBIGUOUS THAI WORDS BY INDUCTIVE LOGIC PROGRAMMING

Miss Chompunuch Kooptiwoot

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Computer Science

Department of Computer Engineering

Graduate School

Chulalongkorn University

Academic Year 1999

ISBN 974-332-872-6

หัวข้อวิทยานิพนธ์	การตัดคำกำกวมในข้อความภาษาไทยด้วยการโปรแกรมตรรกะเชิงอุปนัย
โดย	นางสาว ชมภูนุช คุปติวุฒิ
ภาควิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ ดร. บุญเสริม กิจศิริกุล

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์ ดร. สุชาดา กีระนันทน์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(รองศาสตราจารย์ ดร. วันชัย ริ้วไพบูลย์)

..... อาจารย์ที่ปรึกษา

(อาจารย์ ดร. บุญเสริม กิจศิริกุล)

..... กรรมการ

(อาจารย์ ดร. ยรรยง เต็งอำนวย)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. สมชาย ประสิทธิ์จูตระกูล)

ชมภูนุช คุปติวุฒิ : การตัดคำกำกวมในข้อความภาษาไทยด้วยการโปรแกรมตรรกะเชิงอุปนัย (SEGMENTATION OF AMBIGUOUS THAI WORDS BY INDUCTIVE LOGIC PROGRAMMING)
อ. ที่ปรึกษา : อ. ดร. บุญเสริม กิจศิริกุล; 71 หน้า. ISBN 974-332-872-6.

งานวิทยานิพนธ์ฉบับนี้มีวัตถุประสงค์เพื่อประยุกต์ใช้การโปรแกรมตรรกะเชิงอุปนัยหรือไอแอลพีในการตัดคำกำกวมในข้อความภาษาไทย ระบบไอแอลพีที่เลือกใช้คือระบบ FOIL โดยจะนำไปเปรียบเทียบกับระบบการเรียนรู้แบบประพจน์ โดยระบบการเรียนรู้แบบประพจน์ที่เลือกใช้คือระบบ RIPPER ขั้นตอนการวิจัยเริ่มจากการใช้ FOIL และ RIPPER เรียนรู้คุณลักษณะของคำกำกวม ผลที่ได้จากการเรียนรู้ของระบบ FOIL คือกลุ่มของอนุประโยคฮอร์นอันดับที่หนึ่ง ส่วนผลที่ได้จากการเรียนรู้ของระบบ RIPPER คือกลุ่มของกฎประพจน์ ซึ่งแต่ละอนุประโยคหรือกฎจะนิยามคุณลักษณะของคำกำกวมแต่ละคำ จากการทดลองพบว่ากลุ่มของอนุประโยคที่ได้จากการเรียนรู้ด้วยระบบ FOIL สามารถนิยามคุณลักษณะของคำกำกวมได้ถูกต้องมากกว่ากลุ่มของกฎที่ได้จากการเรียนรู้ด้วยระบบ RIPPER ขั้นตอนถัดมาคือการนำกลุ่มของอนุประโยคที่ได้จากการเรียนรู้ด้วยระบบ FOIL มาช่วยในการตัดคำกำกวมในข้อความภาษาไทย จากการทดลองได้ว่าการตัดคำกำกวมในข้อความภาษาไทยที่มีการใช้กลุ่มของอนุประโยคที่ได้จากการเรียนรู้ด้วยระบบ FOIL ให้ความถูกต้องมากกว่าวิธีการตัดคำแบบจำลองไตรแกรม

ภาควิชา วิศวกรรมคอมพิวเตอร์ ..... **ลายมือชื่อนิสิต** ชมภูนุช คุปติวุฒิ .....
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ..... **ลายมือชื่ออาจารย์ที่ปรึกษา** .....
ปีการศึกษา 2542 ..... **ลายมือชื่ออาจารย์ที่ปรึกษาร่วม** .....

3970382921: MAJOR COMPUTER SCIENCE
KEY WORD:
INDUCTIVE LOGIC PROGRAMMING / PROPOSITIONAL LEARNING SYSTEM / FIRST-ORDERHORN CLAUSE / TRIGRAM MODEL

CHOMPUNUCH KOOPTIWOOT : SEGMENTATION OF AMBIGUOUS THAI WORDS BY INDUCTIVE LOGIC PROGRAMMING. THESIS ADVISOR : BOONSERM KIJSIRIKUL, Ph.D. 71 pp.
ISBN 974-332-872-6.

The purpose of this thesis is to apply Inductive Logic Programming (ILP) to the segmentation of ambiguous Thai words. The ILP system which has been chosen is FOIL. Another learning system, which is used to be compared with FOIL, is a propositional learning system, RIPPER. First, FOIL and RIPPER are used to learn features of ambiguous Thai words. The outputs of FOIL and RIPPER are a set of first-order Horn clauses and a set of propositional rules respectively, each of which defines the features of the ambiguous Thai words. Experimental results show that the clauses learned by FOIL are more accurate than rules of RIPPER. Then, these clauses from FOIL are used to help for segmentation of ambiguous Thai words. Experimental results show that the usage of the clauses improves the accuracy of word segmentation which uses trigram model.

ภาควิชา วิศวกรรมคอมพิวเตอร์	**ลายมือชื่อนิสิต** โฆมปนุช คุปติกุล
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์	**ลายมือชื่ออาจารย์ที่ปรึกษา**
ปีการศึกษา 2542	**ลายมือชื่ออาจารย์ที่ปรึกษาร่วม**

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของอาจารย์ดร.บุญเสริม กิจศิริกุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้ให้คำแนะนำ ข้อคิดเห็นต่าง ๆ ในการวิจัยด้วยดีตลอดมา และตรวจแก้วิทยานิพนธ์ฉบับนี้อย่างละเอียด

ขอขอบคุณ คุณไพศาล เจริญพรสวัสดิ์ ที่ได้ให้ข้อมูลและโปรแกรมการตัดคำแบบจำลองไตรแกรมที่นำมาใช้ในงานวิจัย และขอขอบคุณพี่ๆ และเพื่อนๆ ที่ได้ให้คำปรึกษา กำลังใจและความช่วยเหลือในด้านต่างๆ ซึ่งทำให้การทำงานวิจัยเป็นไปอย่างราบรื่น

ท้ายนี้ ผู้วิจัยใคร่ขอกราบขอบพระคุณ บิดา มารดา ซึ่งสนับสนุนในด้านการเงินและให้กำลังใจแก่ผู้วิจัยเสมอจนสำเร็จการศึกษา

สารบัญ

หน้า

หน้า

บทที่

สารบัญตาราง

หน้า

สารบัญภาพ

หน้า