การแก้คำผิดแบบไม่ตั้งใจโดยอัตโนมัติในภาษาไทย เพื่อการสื่อสารกับหุ่นยนต์สนทนา

นางสาววนิดา เกษร์สุวรรณ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2548

ISBN 974-53-2846-4

AUTOMATIC CORRECTION FOR UNINTENTIONAL TYPING ERRORS IN THAI LANGUAGE FOR COMMUNICATION WITH CHAT ROBOT

Ms. Wanida Kessuwon

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Computer Science
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2005

ISBN 974-53-2846-4

หัวข้อวิทยานิพนธ์	การแก้คำผิดแบบไม่ตั้งใจโดยอัตโนมัติในภาษาไทย เพื่อการสื่อสารกับหุ่นยนต์สนทนา
โดย	นางสาววนิดา เกษร์สุวรรณ์
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ ดร.วิษณุ โคตรจรัส

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.ดิเรก ลาวัณย์ศิริ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ทวิตีย์ เสนีวงศ์ ณ อยุธยา)

..... อาจารย์ที่ปรึกษา
(อาจารย์ ดร.วิษณุ โคตรจรัส)

..... กรรมการ
(อาจารย์ ดร.โปรดปราน บุณยพุกกณะ)

..... กรรมการ
(อาจารย์ ดร.อรรถวิทย์ สุดแสง)

นางสาววนิดา เกษร์สุวรรณ์ : การแก้คำผิดแบบไม่ตั้งใจโดยอัตโนมัติในภาษาไทย เพื่อการสื่อสารกับหุ่นยนต์สนทนา. (AUTOMATIC CORRECTION FOR UNINTENTIONAL TYPING ERRORS IN THAI LANGUAGE FOR COMMUNICATION WITH CHAT ROBOT) อ. ที่ปรึกษา : อาจารย์ ดร.วิษณุ โคตรจรัส, 82 หน้า. ISBN 974-53-2846-4.

วิทยานิพนธ์นี้นำเสนอวิธีแก้ไขคำผิดแบบไม่ตั้งใจโดยอัตโนมัติในภาษาไทย เพื่อนำไปใช้กับหุ่นยนต์สนทนา ซึ่งใช้คุณสมบัติอักขระข้างเคียงบนแป้นพิมพ์ช่วยในการแก้ไขคำผิด โดยได้มีการทดสอบเพื่อสังเกตพฤติกรรมในการพิมพ์ พบรูปแบบการพิมพ์ผิดในภาษาไทยว่า 93.54 เปอร์เซ็นต์จากคำผิดที่พบทั้งหมด เกิดจากความผิดพลาดทั้ง 4 กรณีประกอบกัน คือ แทนที่ เกิน ตก สลับ (เรียงตามลำดับปริมาณที่พบจากมากไปน้อย) และตำแหน่งอักขระที่ผิดนั้นเฉลี่ยอยู่ตำแหน่งที่ 58.36 เปอร์เซ็นต์ของความยาวคำ

ในงานวิจัยนี้นำรูปแบบการพิมพ์ผิดที่ได้มาออกแบบอัลกอริทึมแก้ไขคำผิด แล้วทำการทดสอบประสิทธิภาพด้วยบทสนทนาที่มีคำผิดแบบไม่ตั้งใจทั้งหมดจำนวน 120 ประโยค (ยกเว้นความผิดพลาดที่มาจากการพิมพ์ตก) พบว่าหุ่นยนต์สามารถตอบได้คิดเป็น 95 เปอร์เซ็นต์ ซึ่งแสดงให้เห็นว่าหุ่นยนต์สามารถทำงานได้ประสิทธิภาพมากขึ้นหากใช้อัลกอริทึมนี้

ภาควิชา.... วิศวกรรมคอมพิวเตอร์.....ลายมือชื่อนิสิต.....

สาขาวิชา....วิทยาศาสตร์คอมพิวเตอร์.....ลายมือชื่ออาจารย์ที่ปรึกษา.....

ปีการศึกษา .....2548.....

46704653 : MAJOR Computer Science

KEY WORD: CHAT ROBOT / PATTERN MATCHING / AIML / AUTOMATIC CORRECTION / UNINTENTIONAL TYPING ERRORS / WORD SEGMENTATION

WANIDA KESSUWON : AUTOMATIC CORRECTION FOR UNINTENTIONAL TYPING ERRORS IN THAI LANGUAGE FOR COMMUNICATION WITH CHAT ROBOT. THESIS ADVISOR : VISHNU KOTRAJARAS, Ph.D., 82 pp.

ISBN 974-53-2846-4.

This thesis proposes an algorithm, which is an automatic correction algorithm for unintentional typing errors in Thai language, for communicating with chat robot. It uses the characteristic of adjacent alphabets on keyboard for error correction. Our investigation found that 93.54 percent of all misspelled words in Thai language contains four kinds of typing errors : substitution, insertion, deletion and transposition (arranged from maximum to minimum frequency). Average error position in misspelled words is at 58.36 percent of the word length.

This thesis uses the discovered typing error patterns to design an error correction algorithm. The algorithm is tested on 120 sentences with unintentional typing errors (except error from deletion typing). The chat robot is able to identify 95 percent of the errors. It shows that the chat robot can work more effectively if it uses this algorithm.

Department.... Computer Engineering.... Student's signature Wanida Kessuwon

Field of study.... Computer Science..... Advisor's signature .....

Academic year ...2005.....

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จได้ด้วยความกรุณาของอาจารย์ ดร.วิษณุ โคตรจรัส อาจารย์ที่ปรึกษา ซึ่งเป็นผู้ให้ความรู้ คำปรึกษา และข้อคิดเห็นอันเป็นประโยชน์ต่องานวิจัย รวมทั้งให้โอกาส ความช่วยเหลือ และกำลังใจที่ดีแก่ผู้วิจัยเสมอมา

ขอขอบคุณคณะกรรมการสอบวิทยานิพนธ์ ที่ได้กรุณาให้คำแนะนำในการแก้ไขวิทยานิพนธ์ให้มีคุณภาพยิ่งขึ้น

ขอขอบคุณอาจารย์ ดร.อรรถวิทย์ สุดแสง ที่กรุณาให้ใช้ห้องปฏิบัติการ ISL2 เป็นเวลา 5 เดือน และได้ให้ข้อคิดเห็น และคำแนะนำอันเป็นประโยชน์เกี่ยวกับการหาสถิติการพิมพ์ผิดในภาษาไทยของงานวิจัยนี้

ขอขอบคุณอาจารย์ ดร.โปรดปราน บุณยพุกกณะ และอาจารย์ ดร.อติวงศ์ สุชาโต ที่ให้ความกรุณาแก่ผู้วิจัยได้เป็นส่วนหนึ่งของห้องปฏิบัติการ SLS เป็นเวลา 9 เดือน

ขอขอบคุณอาจารย์ ดร.อรรถสิทธิ์ สุรฤกษ์ ที่ให้ความช่วยเหลือในการประสานงานกับเนคเทคเกี่ยวกับข้อมูลของโปรแกรมตัดคำภาษาไทย Swath และให้คำปรึกษาด้านต่างๆ เสมอมา

ขอขอบคุณอาจารย์ฐานิศรา เกียรติบารมี ที่คอยให้คำปรึกษาในยามที่ข้าพเจ้าท้อแท้ และให้กำลังใจที่ดีเสมอ

ขอขอบคุณอาจารย์ เพื่อนๆ พี่ๆ และน้องๆ ในภาควิชาที่ช่วยทำแบบทดสอบในการเก็บสถิติการพิมพ์ผิดในภาษาไทย และได้ให้ข้อความบทสนทนาบางส่วนของตนเอง เพื่อให้ผู้วิจัยนำมาใช้เป็นตัวอย่างข้อมูลในการออกแบบและทดสอบงานวิจัย

ขอบคุณน้องกวิน อุชุกานนท์ชัย ที่ให้คำปรึกษาและความช่วยเหลือในการเขียนโปรแกรมด้วยภาษาจาวา พร้อมทั้งให้โค้ดตัวอย่างการทำผลคูณคาร์ทีเซียน

ขอบคุณสมาชิกห้องปฏิบัติการชั้น 20 และห้องปฏิบัติการอื่นๆ ที่คอยช่วยเหลือ ให้คำปรึกษา และให้กำลังใจในการทำวิจัยนี้

ท้ายที่สุด ผู้วิจัยใคร่ขอขอบพระคุณ คุณพ่อ คุณแม่ และญาติพี่น้องทุกคนในครอบครัว ที่คอยดูแลห่วงใย เป็นกำลังใจ และให้การสนับสนุนในทุกๆ ด้านตลอดมา

สารบัญ

หน้า

บทที่ หน้า

สารบัญตาราง

ตารางที่ หน้า

สารบัญภาพ

ภาพที่ หน้า