

ชื่อเรื่อง	เครื่องมือตรวจสอบคำที่สะกดผิดในเอกสารภาษาไทย
ผู้วิจัย	นางสาวภัศราภรณ์ สมบูรณ์ศักดิ์
คณะ	วิทยาศาสตร์
ปีที่วิจัย	2554

บทคัดย่อ

การศึกษานี้สร้างเครื่องมือสำหรับตรวจสอบคำสะกดผิดในเอกสารภาษาไทย จากปัญหาของการจัดทำเอกสารในภาษาไทยผิดพลาดมากขึ้น ทำให้ต้องนำเครื่องมือมาใช้ช่วยตรวจสอบคำ กระบวนการตรวจสอบเริ่มจากการสร้างพจนานุกรมคำและยึดตามเกณฑ์ราชบัณฑิตยสถาน หลังจากรวบรวมคำได้ครอบคลุมแล้วทำการตัดคำให้เป็นคำที่เล็กที่สุดที่มีความหมายคงเดิม ตามวิธีการตัดคำแบบกฎ Thai Character clusters (TCC) และนำวิธีการตัดคำแบบสอดคล้องมากที่สุด (Maximal matching) มาใช้ตัดคำร่วมด้วย การตรวจสอบคำทำการตรวจอักขระที่แยกคำได้ทุกตัวอักขระ นำอักขระที่ได้มาคำนวณหาค่าความน่าจะเป็นของคำโดยดูจากความถี่ที่เกิดคำและความน่าจะเป็นในการเกิดร่วมกันของคำในฐานข้อมูล สำหรับตัวอักขระที่ตัดคำแล้วไม่ได้เป็นส่วนหนึ่งของคำ หรือไม่ปรากฏในพจนานุกรมจะถูกตัดทิ้ง กระบวนการแนะนำคำที่ถูกต้องพิจารณาจากบริบทของประโยค และการเกิดร่วมกันของคำ (Feature Based Approach) ผลการนำระบบไปใช้ตรวจสอบเอกสารที่ใช้ภาษาไทยระบบสามารถตรวจสอบอักขระในเอกสารที่ใช้ภาษาไทยอย่างเป็นทางการถูกต้องเฉลี่ยร้อยละ 93.63 และระบบสามารถตรวจสอบอักขระที่ใช้ภาษาไทยอย่างไม่เป็นทางการถูกต้องเฉลี่ยร้อยละ 84.62 จากผลการทดลองที่ได้แสดงว่าสามารถนำไปตรวจสอบเอกสารที่เป็นทางการ และไม่ใชเป็นทางการได้อย่างมีประสิทธิภาพ

คำสำคัญ : สะกดคำผิด การตรวจหาความผิดพลาด การแก้ความผิดพลาด การประมวล
ภาษาธรรมชาติ