BI102-43

ซอฟต์แวร์ตัดคำและกำกับหน้าที่คำภาษาไทยอัตโนมัติ

ระบบตัดคำภาษาไทยและกำกับหน้าที่คำอัตโนมัติ โดยใช้โมเดลทางสถิติซึ่งเป็นคลังข้อความที่ทำการตัดคำและกำกับหน้าที่คำโดยนักภาษาศาสตร์

คุณสมบัติ

1. เป็นระบบตัดคำภาษาไทยและกำกับหน้าที่คำอัตโนมัติโดยใช้โมเดลทางสถิติ เทคนิคที่ใช้ได้แก่ ฮิดเดนมาคอฟโมเดล (Hidden Markov Model) และวิเทอร์บิอัลกอลิทึม (Viterbi Algorithm) สำหรับคลังข้อความที่นำมาใช้ในการหาค่าสถิติต่างๆ คือคลังข้อความออร์คิด ซึ่งเป็นคลังข้อความที่ทำการตัดคำและกำกับหน้าที่คำโดยนักภาษาศาสตร์

2. เป็นผลงานที่อยู่ในรูปแบบต่างๆ ซึ่งนำไปใช้งานได้หลายแบบ เช่น เป็นโปรแกรมที่ทำงานบนดอส บนไมโครซอฟท์วินโดว์ 95 ขึ้นไป บนระบบปฏิบัติการตระกูลยูนิกซ์ ลีนุกซ์ ไลบรารีภาษาซี และไดนามิคลิงค์ไลบรารี (Dynamic Link Library)

3. การตัดคำและกำกับหน้าที่คำอัตโนมัติสามารถนำไปประยุกต์เพื่อพัฒนาเทคโนโลยีต่างๆ ที่เกี่ยวกับการประมวลผลภาษาไทย เช่น การนำไปใช้ร่วมกับเท็กซ์อิดิเทอร์ (Text Editor) เว็บบราวเซอร์ (Web Browser) ระบบอ่านเสียงจากข้อความ (Text to speech) หรือระบบแปลภาษาไทยเป็นภาษาอังกฤษ เป็นต้น

ประโยชน์ที่ได้รับ

เป็นพื้นฐานในการพัฒนาเทคโนโลยีของซอฟต์แวร์ประมวลผลภาษาไทย ทำให้ได้ซอฟต์แวร์ที่มีคุณสมบัติดี

วิจัยและพัฒนาโดย: งานวิจัยและพัฒนาสาขาสารสนเทศซอฟต์แวร์ ภาษาและวิศวกรรมความรู้
โครงการพิเศษและพัฒนาการตัดคำภาษาไทยและกำกับหน้าที่คำอัตโนมัติ
สอบถามข้อมูลทางเทคนิคเพิ่มเติมได้ที่หน้า 249

รหัสผลงาน: BI102-43

ติดต่อสอบถามได้ที่: ฝ่ายพัฒนาธุรกิจและอุตสาหกรรม ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ
โทร. (662) 644-8150..99 ต่อ 610, 656..9
โทรสาร (662) 644-8122, 644-8137..8