

หัวข้อวิทยานิพนธ์	การตัดคำไทยโดยใช้ดิคชันนารีที่มีโครงสร้างข้อมูลแบบแฮชซิง
นักศึกษา	นายสรศักดิ์ ไทยแท้
อาจารย์ผู้ควบคุมวิทยานิพนธ์	ผศ.ดร.บุญธีร์ เครือตราฐ
หลักสูตร	วิทยาศาสตร์มหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร ลาดกระบัง
พ.ศ.	2541

### บทคัดย่อ

เนื่องจากตัวอักษรไทยมักจะเขียนติดกันไม่มีตัวอักษรพิเศษขึ้นระหว่างคำดังในภาษาอังกฤษ จึงเป็นการยากที่จะทราบว่าผู้เขียนต้องการจะแบ่งคำอย่างไร หรือ คำใดมีคำที่สะกดผิดอยู่ ทั้งนี้การรู้ขอบเขตของคำโดยอัตโนมัติ จะช่วยให้สามารถตรวจสอบคำสะกดได้ถูกต้องและรวดเร็วขึ้นทั้งสามารถทำ index จากคำในข้อความได้ งานวิจัยนี้ได้นำเสนออัลกอริทึมในการสืบค้นคำศัพท์และพยายามตัดแบ่งคำโดยอัตโนมัติโดยใช้พจนานุกรม (Dictionary) ที่มีโครงสร้างแบบแฮชซิง (Hashing) เป็นตัวช่วยในการตัดคำ รวมทั้งได้ออกแบบอัลกอริทึมแบบ Iterative และโครงสร้างข้อมูล Stack ที่ใช้ในการเก็บตำแหน่งคำในการตัดคำแทนแบบเดิมที่เป็น Recursive รวมทั้งพจนานุกรมที่ใช้แบบ Hashing ทำให้สามารถเพิ่มเติมคำได้ง่ายรวมทั้งสามารถเลือกหาคำคล้ายได้สะดวก งานวิจัยนี้ยังได้ทำการเปรียบเทียบการตัดคำและการตรวจสอบคำผิดกับโปรแกรม Commercial คือ Word 6.0 และ Word 7.0 ของบริษัทไมโครซอฟต์รวมทั้งได้รวบรวม ข้อผิดพลาดของ Word 6.0 และ Word 7.0 มาทำการปรับปรุงอัลกอริทึมในงานวิจัยนี้ งานวิจัยนี้มีข้อดีคือสามารถตัดคำได้ทุกตำแหน่ง ถ้าคำนั้นมีในพจนานุกรมซึ่งจะพบคำในประโยคที่ถูกต้องตามพจนานุกรมและยังใช้ความยาวมากที่สุดของคำไทยในพจนานุกรม (25 ตัวอักษร) มาช่วยในการเพิ่มความเร็วในการตัดคำโดยทำการพิจารณาคำที่ละ 25 ตัวอักษร นอกจากนี้ยังเพิ่มเติมโครงสร้างพจนานุกรมนี้ให้มีการหยุดตัดคำที่สั้นกว่า 25 ตัวอักษรเมื่อพบว่าคำนั้นไม่มีตัวอักษรต่อในพจนานุกรม แต่มีข้อเสียคือในการค้นหาคำจะช้ากว่าโครงสร้างข้อมูลแบบทรี (Trie)