บทที่ 1
บทนำ

1.1 ความเป็นมาของปัญหา

1.1.1 ลักษณะของภาษาไทยและปัญหา

ปัจจุบันเทคโนโลยีทางคอมพิวเตอร์ ได้พัฒนาระบบไมโครคอมพิวเตอร์ให้มีประสิทธิภาพมากขึ้นจนสามารถนำไปใช้งานทางธุรกิจได้ การประยุกต์ในงานที่สำคัญมากอันหนึ่ง คือ การนำไมโครคอมพิวเตอร์ไปใช้พิมพ์เอกสารแทนเครื่องพิมพ์ดีดธรรมดา หรือที่เรียกว่าการประมวลผลคำ (word processing) ได้ช่วยให้ งานผลิตเอกสารที่เป็นสิ่งจำเป็นอย่างมากในระบบการบริหารและเชิงธุรกิจอื่นๆ มีประสิทธิภาพและมีความรวดเร็วขึ้น แต่การทำงานของการประมวลผลคำภาษาไทยยังมีอุปสรรคบางประการอันเนื่องมาจากลักษณะพื้นฐานของภาษาไทยเองคือหน่วยคำ หรือ พยางค์ในภาษาไทยจะเรียงติดต่อกันโดยไม่มีเครื่องหมายพิเศษหรือช่องว่างคั่นกลางระหว่างหน่วยคำ เหมือนกับประโยคในภาษาอังกฤษ มีเพียงช่องว่างที่ใช้เพื่อการแบ่งแยกระหว่างประโยค หรือ วลี หรือ เพื่อแยกวรรคตอนขึ้นย่อหน้าใหม่เท่านั้น การที่ไม่มีเครื่องหมายการแบ่งแยกหน่วยคำนี้ทำให้เกิดความยุ่งยากขึ้นอย่างมากกับการประมวลผลข้อมูลภาษาไทย ด้วยคอมพิวเตอร์

ด้วยเหตุนี้การวิจัย เพื่อให้ได้ระบบการตัดคำภาษาไทยที่มีประสิทธิภาพและ เหมาะสมต่อการใช้งาน จึงเป็นสิ่งจำเป็นอย่างยิ่งสำหรับวงการคอมพิวเตอร์ของประเทศไทย โดยมิได้จำกัดอยู่แต่เพียงการประมวลผลคำแต่เพียงอย่างเดียว

1.1.2 การตัดคำ

การตัดแบ่งคำภาษาไทย มีวิธีในการแบ่งคำอยู่ 2 วิธี คือ

1.1.2.1 การตัดคำโดยใช้กฎ (Rule)

1.1.2.2 การตัดคำโดยใช้พจนานุกรม (Dictionary)

ในที่นี้พิจารณาเฉพาะการตัดคำโดยใช้พจนานุกรม ซึ่งจะให้ความถูกต้องในการตัดคำสูง ปัจจุบันได้มีวิธีการตัดคำโดยใช้พจนานุกรม โดยวิธีการดังกล่าวจะนำพจนานุกรมเข้ามาเก็บไว้ในหน่วยความจำหลักทั้งหมด แต่มีข้อเสียคือเมื่อคำศัพท์มีจำนวนมากขึ้น จะทำให้หน่วยความจำหลักของเครื่องเต็มไม่สามารถเพิ่มคำศัพท์ได้อีก ดังนั้นจึงควรมีการแก้ไขโครงสร้างของพจนานุกรมให้สามารถแบ่งเป็นส่วนได้ เมื่อจะนำพจนานุกรมเข้าสู่หน่วยความจำหลักก็นำขึ้นมาเพียงบางส่วน

1.2 วัตถุประสงค์ของการวิจัย

การทำระบบพจนานุกรมเสมือนเพื่อให้สามารถใช้ระบบการตัดคำกับพจนานุกรมที่มีขนาดใหญ่กว่าขนาดของหน่วยความจำหลักได้

1.3 ขอบเขตของการวิจัย

1. พัฒนาระบบบน ไมโครคอมพิวเตอร์ ภายใต้ระบบปฏิบัติการแบบ MS-DOS ซึ่งไม่มีระบบหน่วยความจำเสมือนอยู่ก่อน
2. ในการวิจัยจะใช้วิธีการตัดคำที่มีอยู่แล้ว (สัมพันธ์ , 1991)
3. รหัสภาษาไทยที่ใช้ จะใช้รหัสของสำนักงานมาตรฐานผลิตภัณฑ์อุตสาหกรรม(ส.ม.อ.)
4. การพัฒนาโครงสร้างของพจนานุกรม และการสร้างพจนานุกรมเสมือนจะอยู่ภายใต้ระบบภาษาซี (C language)

1.4 ขั้นตอนและวิธีดำเนินงานวิจัย

1. ออกแบบโครงสร้างของพจนานุกรม ให้สามารถแบ่งเป็นส่วนๆได้
2. พัฒนาโปรแกรมสร้างพจนานุกรมเสมือน ที่สามารถดึง คำศัพท์ที่ต้องการที่อยู่ในหน่วยความจำสำรองขึ้นมาอยู่ในหน่วยความจำหลัก และเลือกส่วนของพจนานุกรมที่เดิมอยู่ในหน่วยความจำหลักออกเพื่อนำส่วนที่ต้องการใช้เข้ามาแทนที่

3. สร้างตารางแสดงความสัมพันธ์ระหว่างแต่ละตัวอักษรในคำกับตำแหน่งของตัวอักษรนั้นๆในหน่วยความจำเสมือน

4. ทดสอบ, ปรับปรุงแก้ไขโปรแกรมสร้างพจนานุกรมเสมือน

5. สรุปผลการวิจัย

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ทำให้ระบบพจนานุกรมสามารถบรรจุคำศัพท์ได้จำนวนมากขึ้น เป็นผลให้การตัดคำมีความถูกต้องสูงขึ้น

2. ลักษณะของการตัดแบ่งพจนานุกรมเสมือน สามารถนำไปประยุกต์กับโครงสร้างข้อมูลที่มีขนาดใหญ่ในลักษณะเดียวกัน เพื่อให้สามารถใช้งานบนเครื่องไมโครคอมพิวเตอร์ ที่มีหน่วยความจำไม่มากนัก