



บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

นับแต่ได้มีการนำคอมพิวเตอร์เข้าช่วยปฏิบัติงานด้านต่าง ๆ จากอดีต จนถึงปัจจุบัน ไม่ว่าจะเป็นหน่วยงานของรัฐ หรือของเอกชน ระบบงานที่ใช้คอมพิวเตอร์เข้าช่วย ส่วนใหญ่ มีความต้องการใช้วิธีประมวลผลด้วยอักษรภาษาไทย ซึ่งแทบจะกล่าวได้ว่าเกือบทั้งหมด จากลักษณะเฉพาะตัวของภาษาไทยที่แตกต่างกับภาษาอื่น ๆ เช่น มีวรรณยุกต์ หรือ ระดับเสียง (TONE) การพิมพ์มี 4 ระดับ หรือท้องพิมพ์ 4 บรรทัด และการจับคำไม่มีการเว้นวรรคอย่างภาษาอื่น สิ่งเหล่านี้ทำให้เกิดปัญหาในการใช้คอมพิวเตอร์กับภาษาไทย ไม่น้อย นักคอมพิวเตอร์ของไทย ก็พยายามคิดหาวิธีการต่าง ๆ มาเสริมเข้ากับระบบคอมพิวเตอร์ เพื่อให้ระบบคอมพิวเตอร์สามารถปฏิบัติการประมวลผลกับลักษณะของภาษาไทย ได้ตามต้องการ

ในการค้นหาข้อมูล (SEARCHING) จากแฟ้มข้อมูล (FILE) เริ่มต้นจากการใช้รหัสหมายเลขแทนชื่อในการค้นหา การใช้ตัวสะกดภาษาอังกฤษ เป็น ตัวสะกดแทนอักษรภาษาไทย ในการค้นหา มาจนกระทั่งได้มีการออกแบบรหัสสำหรับอักษรภาษาไทย ตัวพิมพ์ และแป้นพิมพ์ภาษาไทย จึงได้มีการใช้รหัสของอักษรภาษาไทยเป็นรหัสค้นหา จากวิธีการต่าง ๆ ที่กล่าวมาข้างต้น จะเห็นว่า การใช้รหัสหมายเลขแทนชื่อในการค้นหา จะต้องจำหรือกำหนดคีย์เลขหรืออักษรทุกตัวที่ใช้เป็นรหัสในการค้นหาให้ถูกต้อง จึงจะค้นหาข้อมูลพบ และเช่นกัน ถ้าใช้ตัวสะกดของชื่อเป็นรหัสในการค้นหา ไม่ว่าจะใช้อักษรภาษาอังกฤษแทน หรือใช้อักษรภาษาไทยโดยตรงก็ตาม ทุกตัวอักษรที่จะใช้จะเป็นรหัสในการค้นหาข้อมูลในแฟ้มข้อมูล ต้องใช้ให้ถูกต้อง และ ตรงกับตัวอักษรของรหัสในแฟ้มข้อมูล จึงจะพบข้อมูลที่ต้องการ ค้นหา การใช้อักษรผิดไปเพียงตัวใดตัวหนึ่ง เช่น พยัญชนะ ค ในภาษาไทยสามารถ ใช้ พยัญชนะภาษาอังกฤษแทนได้หลายตัว คือ K CH หรือ C ถ้าเลือกตัวอักษร

แทนนิก การค้นหาข้อมูลตามรหัสนั้นก็จะมีผล หรืออีกตัวอย่างหนึ่ง คือ การใช้อักษรภาษาไทยโดยตรง ถ้าใช้นิก เช่น น และ ณ หรือ ด และ ร มีเสียงสะกดเหมือนกัน ถ้าใช้ตัวที่ต่างจากที่เก็บไว้ในแฟ้มข้อมูลแล้ว ก็จะทำให้ไม่พบข้อมูลในแฟ้มข้อมูลได้ เช่นเดียวกัน

จากปัญหาของพยัญชนะพ้องเสียง หรือใช้สะกดแทนกันได้ นี้ ในการประมวลผลข้อมูลภาษาอังกฤษ ใ้มีการคิดวิธีเพื่อช่วยให้การค้นหาข้อมูลไม่เกิดการผิดพลาดมาก วิธีการดังกล่าวนี้ เรียกว่า ซาวด์เด็กซ์ (SOUNDEX) วิธีนี้เป็นการจัดกลุ่มพยัญชนะที่มีเสียงสะกดเดียวกันไว้เป็นกลุ่ม ๆ การสร้างรหัสสำหรับค้นหา จะใช้รหัสตัวเดียวกันแทนพยัญชนะในแต่ละกลุ่ม ซึ่งวิธีการดังกล่าว สามารถนำมาประยุกต์ใช้กับภาษาไทยได้ จากการที่พยัญชนะของไทยมีเสียงสะกดเดียวกันอยู่จำนวนมากก็ต้องพิจารณาว่าจะจัดกลุ่มแบบใด จึงจะทำให้การค้นหาข้อมูลในแฟ้มข้อมูล เป็นไปอย่างถูกต้องมากที่สุด

ดังนั้น การวิจัยนี้จะพิจารณาถึงการจัดกลุ่มอักษรภาษาไทยแบบต่าง ๆ เพื่อหาแบบที่ดีที่สุด มาเป็นแบบในการสร้างรหัสสำหรับแฟ้มข้อมูลที่จะใช้ ชื่อ หรือ พยัญชนะ เป็นรหัสค้นหา และจะมีการแปลง ชื่อ หรือ พยัญชนะให้เป็นรหัสตามแบบของการจัดกลุ่มที่สร้างไว้

1.2 วัตถุประสงค์ของการวิจัย

ในการทำวิจัยเรื่องนี้ มีวัตถุประสงค์ดังต่อไปนี้

- 1) เพื่อศึกษาถึงวิธีการจัดกลุ่มอักษรภาษาไทย ว่าจะจัดกลุ่ม และกำหนดรหัสอย่างไร จึงจะทำให้การค้นหาข้อมูลมี ประสิทธิภาพมากที่สุด
- 2) เพื่อหาขนาดความยาวของรหัสจากการจัดกลุ่มว่า ขนาดความยาวของรหัสเท่าไร จะเหมาะสมกับการใช้งานมากที่สุด
- 3) เพื่อศึกษาวิธีจัดแฟ้มข้อมูลแบบต่าง ๆ ว่าแฟ้มข้อมูลแบบใดจะเหมาะสมกับวิธีกำหนดรหัสแบบ ซาวด์เด็กซ์

1.3 ขอบเขตของการวิจัย

ขอบเขตของการวิจัย มีดังนี้

- 1) ทาวิธีจักกลุ่มอักษรภาษาไทย และกำหนดหลักเกณฑ์การให้รหัส ใน
ชาวักเด็กซ์
- 2) พิจารณาการจักเพิ่มข้อมูลว่าจะจักแบบใดจึงจะเหมาะสม และใช้งานได้ดี
- 3) การวิจัยนี้จะคำนึงถึงเฉพาะชื่อคน เท่านั้น ไม่คำนึงถึง ชื่อสิ่งของ
และ อื่น ๆ
- 4) วิเคราะห์และทดสอบประสิทธิภาพ การค้นหาข้อมูลในเพิ่มข้อมูลที่กำหนด
รหัสแบบชาวักเด็กซ์ ว่าการจักกลุ่มและขนาดรหัสอย่างไร จึงจะให้ประสิทธิภาพในการค้นหา
ข้อมูลมากที่สุด

1.4 ขั้นตอนในการทำวิจัย

ขั้นตอนในการดำเนินการวิจัยมีดังนี้

- 1) ศึกษาวิธีการของชาวักเด็กซ์ ในข้อมูลที่เป็นภาษาอังกฤษ ตลอดจนวิธีการ
จักเพิ่มข้อมูล จากหนังสือ วารสาร และเอกสาร ต่าง ๆ
- 2) ศึกษาลักษณะภาษาไทย และ กลุ่มเสียงในอักษรภาษาไทย
- 3) ประยุกต์วิธีการชาวักเด็กซ์ใช้กับภาษาไทย โดยกำหนดหลักเกณฑ์การให้
รหัส ตลอดจนวิธีการสร้างรหัสและเพิ่มข้อมูล
- 4) สร้างโปรแกรมเพื่อทดสอบวิธีการชาวักเด็กซ์ ที่ประยุกต์ใช้ในภาษาไทย
- 5) หาสถิติจากขนาดความยาวของรหัสต่าง ๆ ว่า ความยาวของรหัส
เท่าไร จะเหมาะสมและให้ประสิทธิภาพในการค้นหาข้อมูลที่สุด
- 6) สรุปผล และ ข้อเสนอแนะ

1.5 ประโยชน์ที่ได้รับจากการวิจัย

- 1) เพื่อแนะนำวิธีประยุกต์การใช้ วรรณคดีกับภาษาไทยให้แพร่หลายยิ่งขึ้น
- 2) เพื่อเป็นแนวทางแก่นักศึกษา หรือ นักคอมพิวเตอร์ ในการศึกษาวิธีการจัดกลุ่มตัวอักษรที่มีเสียงพ้องกัน เพื่อพัฒนาการใช้วิธีวรรณคดีต่อไป
- 3) เพื่อประยุกต์กับงานต่าง ๆ ที่ใช้ชื่อเป็นรหัสค้นหา อาทิ งานธนาคาร งานห้องสมุด งานทะเบียนต่าง ๆ เช่น บัตรประชาชน หรืองานที่กิน เป็นต้น