

บทที่ 2

หลักการและทฤษฎีการประมวลผลโดยใช้คำสั่งเสียง

ในการศึกษาพัฒนาการใช้คำสั่งเสียงเพื่อสร้างรายงานจากฐานข้อมูล มีความรู้พื้นฐานที่จำเป็นต้องใช้ในการพัฒนา คือ การเรียนรู้ทำความเข้าใจกับวิธีการนำเสียงพูดเพื่อนำมาใช้เป็นวิธีนำเข้าสู่ข้อมูล และการใช้คำสั่ง เอส คิว แอล เบื้องต้นเพื่อใช้ติดต่อ ดึงข้อมูลจากฐานข้อมูล

2.1 การประมวลผลภาษาธรรมชาติและเทคโนโลยีเสียง

กิตติ ภักดีวัฒน์กุล(2547) ในอดีตจนถึงปัจจุบันผู้ใช้งานยังคงสั่งการคอมพิวเตอร์ด้วยการพิมพ์คำสั่งผ่านทางอุปกรณ์นำเข้าสู่ข้อมูลที่เรียกว่า “คีย์บอร์ด (Keyboard)” หรือคลิกเมนูคำสั่งผ่านทางอุปกรณ์นำเข้าสู่ข้อมูลอีกชนิดหนึ่งซึ่งเรียกว่า “เมาส์ (Mouse)” เมื่อคอมพิวเตอร์ได้รับคำสั่งจากผู้ใช้งานก็จะประมวลผลแล้วได้ตอบกลับผ่านทางอุปกรณ์แสดงผลที่เรียกว่า “จอภาพ (Monitor)” ด้วยรูปแบบการแสดงผลที่หลากหลาย เช่น รูปภาพ สัญลักษณ์ เสียง ข้อความ ภาพเคลื่อนไหว เป็นต้น สำหรับผู้ใช้งานที่ไม่มีความชำนาญทางคอมพิวเตอร์ จะรู้สึกว่าการใช้งานเป็นไปด้วยความยากลำบาก ดังนั้น จึงมีผู้พัฒนาเทคโนโลยีที่มีความฉลาด (Intelligent) ในการเข้าใจคำสั่งจากผู้ใช้งานด้วยเสียงพูด นั่นคือ “การประมวลผลภาษาธรรมชาติ (Natural Language Processing)”

การประมวลผลภาษาธรรมชาติ เป็นการติดต่อสื่อสารระหว่างผู้ใช้งานกับเครื่องคอมพิวเตอร์ โดยที่ผู้ใช้งานสามารถนำเข้าสู่เสียงพูด ซึ่งอาจจะเป็นภาษาอังกฤษ หรือภาษาใดก็ได้ตามที่ผู้พัฒนาได้โปรแกรมเอาไว้ การที่เครื่องคอมพิวเตอร์จะสามารถเข้าใจคำสั่งที่เป็นเสียงได้ คอมพิวเตอร์จะต้องมีองค์ความรู้จัดเก็บไว้เพื่อนำมาวิเคราะห์และแปลเสียงที่นำเข้าสู่เครื่องได้ (การทำความเข้าใจในภาษา ธรรมชาติ) เครื่องคอมพิวเตอร์จึงจะสามารถโต้ตอบกลับไปยังผู้ใช้งานได้ (การสร้างภาษาธรรมชาติ) ในที่นี้จะแบ่งประเภทของระบบประมวลผลภาษาธรรมชาติออกเป็น 2 ชนิด ได้แก่

1) การเข้าใจภาษาธรรมชาติ (Natural Language Understanding)

เป็นขบวนการที่จะทำให้คอมพิวเตอร์สามารถสำรวจหรือวิเคราะห์คำสั่งจากผู้ใช้งานไม่ว่าจะเป็นคำสั่งที่นำเข้าสู่ผ่านคีย์บอร์ดหรือเป็นคำสั่งด้วยเสียง ทั้งนี้ เพื่อสร้างความเข้าใจในคำสั่งนั้น โดยจะแบ่งการทำงานเป็น 2 ส่วน ได้แก่

- การจดจำเสียง (Speech/Voice Recognition) เป็นส่วนการทำงานส่วนแรกเริ่มต้นเมื่อผู้ใช้ป้อนคำสั่งเข้าสู่ระบบหรือเข้าสู่ เครื่องคอมพิวเตอร์ จากนั้นระบบจะทำการจดจำพร้อมกับจำแนกคำพูดที่รับเข้ามาโดยไม่จำเป็นต้องทราบถึงความหมายของคำเหล่านั้น

- การเข้าใจเสียง (Speech/Voice Understanding) เมื่อระบบสามารถจดจำและจำแนกคำสั่งจากผู้ใช้ได้แล้ว ส่วนการทำงานต่อไปคือการสร้างความเข้าใจในคำสั่งที่ได้จดจำและจำแนกไว้ กล่าวคือ จะต้องทราบความหมายของคำเหล่านั้นเอง

2) การสร้างภาษาธรรมชาติ (Natural Language Generation)

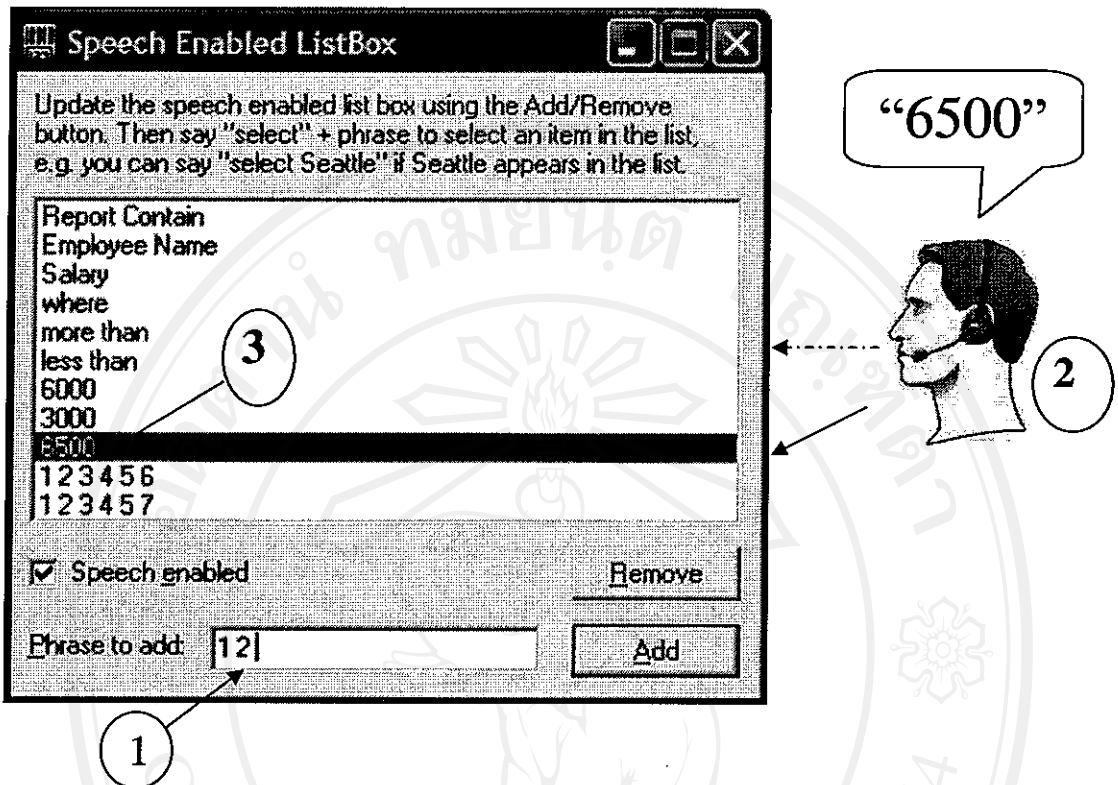
เป็นกระบวนการที่จะทำให้คอมพิวเตอร์สามารถสร้างภาษาธรรมชาติได้เอง เพื่อได้ตอบกลับไปยังผู้ใช้โดยผ่านทางหน้าจอ (Monitor) หรือได้ตอบด้วยเสียง หรือเรียกว่า “กระบวนการสังเคราะห์เสียง (Voice Synthesis)”

- กระบวนการสังเคราะห์เสียง เป็นเทคโนโลยีการสร้างเสียงที่เปล่งออกมาเป็นคำหรือวลี ให้กลายเป็นเสียงอิเล็กทรอนิกส์ (เสียงสังเคราะห์) ที่ได้มีการกำหนดรูปแบบพื้นฐานเสียงอิเล็กทรอนิกส์ไว้แล้ว ปัจจุบัน ได้มีการพัฒนาเสียงอิเล็กทรอนิกส์ให้มีคุณภาพดีขึ้นอย่างต่อเนื่อง แต่เทคโนโลยีการสังเคราะห์เสียงยังมีต้นทุนที่สูงมาก

จะเห็นว่าการประมวลผลภาษาธรรมชาติเป็นการพัฒนาให้คอมพิวเตอร์สามารถโต้ตอบกับมนุษย์ได้ด้วยเสียงพูด ซึ่งนับว่าเป็นความฉลาดที่ทัดเทียมมนุษย์ในอีกด้านหนึ่ง อย่างไรก็ตาม ข้อจำกัดของความชาญฉลาดในด้านการพูดได้ตอบและทำความเข้าใจในภาษาธรรมชาติหรือภาษามนุษย์ยังมีอยู่มาก ข้อจำกัดที่สำคัญที่สุดคือ เครื่องคอมพิวเตอร์หรือระบบประมวลผลภาษาธรรมชาติสามารถจดจำศัพท์และความหมายได้ในจำนวนจำกัด

2.2 โปรแกรมที่เกี่ยวข้องที่ได้ทำการศึกษา

โปรแกรม สปีช อีเนเบิล ลิสต์บ็อกซ์ เป็น โปรแกรมที่ได้ศึกษาและนำมาใช้ในการค้นคว้า โดยเป็น โปรแกรมตัวอย่างที่มีอยู่ใน ไมโครซอฟท์ สปีช เอสดีเค 5.1 โปรแกรมนี้เป็น โปรแกรมที่มีความสามารถในการตรวจสอบเสียง เพื่อนำไปเทียบกับรายการที่ปรากฏ และทำการเลือกรายการความเสียงที่เรานำเข้าผ่านทางไมโคร โฟน ดังรูป 2.1



รูป 2.1 แสดงขั้นตอนการทำงานของโปรแกรม สปีช อีเนเบิล ลิสต์บ็อกซ์

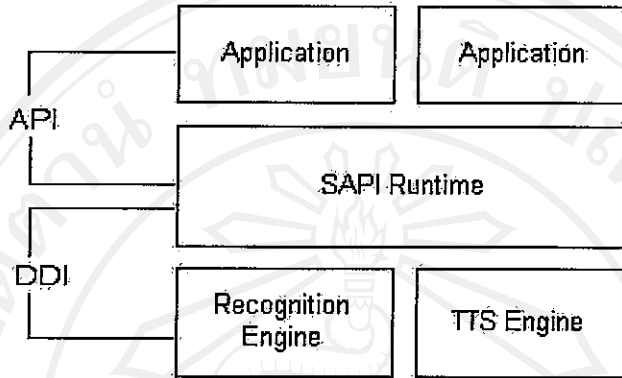
จากรูป 2.1 แสดงขั้นตอนการทำงานของโปรแกรม สปีช อีเนเบิล ลิสต์บ็อกซ์ โดย ขั้นตอนการทำงานของโปรแกรม ประกอบด้วย 3 ขั้นตอนหลัก คือ

- 1) การเพิ่มชุดคำสั่งเข้าไปในรายการ
- 2) การทดสอบการเลือกรายการ โดยการพูดชุดคำสั่งที่ได้เพิ่มไว้ในรายการตามข้อ 1
- 3) โปรแกรมจะทำการเทียบคำพูดกับ คำที่มีในรายการ เมื่อพบคำที่ตรงกัน จะทำการเลือกรายการให้โดยทำแถบสีที่รายการนั้นๆ

2.3 สถาปัตยกรรมภายในของโปรแกรม สปีช อีเนเบิล ลิสต์บ็อกซ์

โปรแกรม สปีช อีเนเบิล ลิสต์บ็อกซ์ ทำงานร่วมกับส่วนติดต่อที่เรียกว่า SAPI (Speech Application Programming Interface) โดยจะแบ่งการทำงานเป็น 2 ส่วน คือ API (Application Programming interface) จะทำงานด้าน high-level รองรับการติดต่อระหว่าง Application กับ Speech Engine และ DDI (Device Driver Interface) จะทำงานทางด้าน low-level เพื่อให้ Engine-level สามารถติดต่อกับ Application ได้ เช่นสามารถรับข้อมูลเข้าผ่านไมโครโฟน เมื่อ Engine-level ตรวจสอบได้ ก็สามารถทำให้เกิดเหตุการณ์เพื่อสั่งให้โปรแกรมทำงานได้ โดยที่

Engine-level ยังแบ่งออกเป็นสองส่วนได้แก่ Recognition Engine ทำหน้าที่ในการรู้จำ และ TTS Engine ทำหน้าที่ในการแปลงตัวอักษรออกมาเป็นคำพูด ดังแสดงในรูป 2.2



รูป 2.2 แสดงสถาปัตยกรรมของ SAPI

ในการทำงานของโปรแกรม เริ่มต้นโปรแกรมจะทำงานผ่าน API โดยตัวโปรแกรมจะทำงานผ่านทาง window application ในการรับข้อมูลตัวอักษร โดยป้อนตัวอักษรผ่านทาง text box นำไปเพิ่มไว้ในรายการ เมื่อมีการพูดผ่านไมโครโฟน Application จะทำงานร่วมกับ Recognition Engine ในการรับข้อมูลเสียง เพื่อนำมาเปรียบเทียบกับข้อมูลที่เราป้อนผ่าน text box ในตอนแรก และเปรียบเทียบค่าเมื่อได้ค่าตรงตามต้องการ โปรแกรมก็จะทำการเลือกรายการนั้น