บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การค้นคืนข้อมูลจากระบบฐานข้อมูลผู้ใช้ระบบต้องทำการศึกษาถึงวิธีการใช้คำสั่งสอบถามต่างๆอย่างละเอียด เนื่องจากไม่มีระบบฟรอนต์เอนด์เพื่ออำนวยความสะดวกในการใช้งานแต่อย่างใด ในยุคถัดมาผู้ผลิตชุดคำสั่งสำเร็จรูปได้ให้ความสนใจที่จะพัฒนาตัวเชื่อมประสานกับผู้ใช้ให้สามารถใช้งานได้ง่ายขึ้นโดยใช้วิธีการต่างๆ เช่น ระบบวินโดวส์ ระบบรายการคำสั่ง เป็นต้น เพื่อใช้แทนชุดคำสั่งสอบถามที่พัฒนาขึ้นมาเพื่อใช้งานกับระบบฐานข้อมูลนั้นๆ อย่างไรก็ตามแนวความคิดในลักษณะเช่นนี้ยังมีข้อจำกัดอยู่ เช่น

1. การตีความ ความหมายของสัญรูป (Icon) ในระบบวินโดว์ หรือการใช้คำสั่งสั้นๆในระบบรายการคำสั่ง อาจทำให้ผู้ใช้ระบบฐานข้อมูลเกิดความเข้าใจผิด หรือไม่มีความสะดวกในการใช้ ซึ่งสาเหตุเกิดจากการออกแบบสัญรูปหรือคำสั่งนั้นไม่สามารถสื่อความหมายตรงกับคำสั่งสอบถามหรือให้ความหมายที่คลุมเครือไม่ชัดเจน

2. ความซับซ้อน การใช้คำสั่งในแต่ละชุดคำสั่งอาจมีความยุ่งยากหรือซับซ้อนมากจนผู้ใช้ที่ไม่คุ้นเคยหรือไม่เข้าใจการใช้งานระบบฐานข้อมูลไม่สามารถใช้งานได้สะดวก ทำให้ผู้ใช้ต้องเสียเวลากับการศึกษาระบบให้เข้าใจเสียก่อน ทั้งที่ในความเป็นจริงผู้ใช้ไม่ต้องการทราบการทำงานของระบบฐานข้อมูลอย่างลึกซึ้งมากนัก

3. ความยืดหยุ่นในการออกรูปแบบรายงาน ในบางระบบไม่สามารถที่จะนำเสนอรายงานในรูปแบบที่แตกต่างไปจากที่ทางหน่วยพัฒนาระบบหรือหน่วยเอ็มไอเอส (MIS – Management Information System) จัดทำไว้ให้ ดังนั้นถ้าหากต้องการรูปแบบรายงานใหม่ๆจะต้องให้หน่วยงานที่พัฒนาระบบจัดการให้ ซึ่งทำให้ต้องเสียเวลาในการจัดทำขึ้นมาทั้งๆที่รูปแบบรายงานนั้นอาจใช้เพียงครั้งเดียว

ถึงแม้ว่าตัวเชื่อมประสานที่มีเพิ่มขึ้นมาเหล่านี้จะทำให้ผู้ใช้สามารถใช้งานได้ง่ายมากขึ้นก็ตาม แต่ผู้ใช้ยังคงต้องอาศัยเวลาเพื่อทำการเรียนรู้เมื่อมีการเปลี่ยนแปลงระบบใหม่ เนื่องจากตัวเชื่อมประสานของชุดคำสั่งสำเร็จรูปแต่ละตัวจะมีลักษณะที่แตกต่างกัน โดยที่ผู้ผลิตมีความพยายามที่จะออกแบบตัวเชื่อมประสานเหล่านี้ให้มีลักษณะใช้งานง่ายมากที่สุด มีผลให้เกิดรูปแบบของตัวเชื่อมประสานระบบฐานข้อมูลมากมายในท้องตลาด

ในปัจจุบันได้เริ่มมีการนำความรู้การประมวลผลภาษาธรรมชาติหรือเอ็นแอลพี (NLP - Natural Language Processing) มาใช้กับระบบฟรอนต์เอนด์เพื่อให้ผู้ใช้สามารถใช้งานได้ง่ายมากขึ้นโดยไม่จำเป็นต้องเรียนรู้เพิ่มเติม และเป็นการลดเวลาและภาระในการพัฒนาระบบฐานข้อมูลของหน่วยพัฒนาอีกด้วย โดยระบบฟรอนต์เอนด์ดังกล่าวจะรับข้อความจากแป้นพิมพ์หรือเสียงพูดในการสั่งงาน นอกจากนี้วิธีการดังกล่าวทำให้ผู้ใช้ระบบฐานข้อมูลมีความสามารถในการทำงานได้มากกว่าโดยใช้เวลาในการทำการศึกษาระบบฐานข้อมูลเพียงเล็กน้อย เนื่องจากระบบดังกล่าวเพียงแค่รับคำสั่งที่เป็นการขอค้นคืนข้อมูลจากผู้ใช้ได้หลายหลากรูปแบบตามแต่ผู้ใช้ถนัด โดยแปลประโยคเหล่านั้นออกมาเป็นคำสั่งสอบถามของแต่ละระบบฐานข้อมูลที่ใช้อยู่ได้ และผู้ใช้ยังสามารถกำหนดรูปแบบรายงานต่างๆได้ตามความต้องการโดยไม่จำเป็นต้องทำการปรับเปลี่ยนหรือแก้ไขระบบอีกด้วย มีผลทำให้องค์กรที่ดูแลระบบฐานข้อมูลสามารถลดค่าใช้จ่ายและเวลาในการพัฒนาระบบได้อย่างมาก

ตัวอย่างของชุดคำสั่งสำเร็จรูปที่นำเอ็นแอลพีในระบบฟรอนต์เอนด์ได้แก่ ชุดคำสั่งสำเร็จรูปที่ชื่อ Natural Language ของ Natural Language Inc. เป็นชุดคำสั่งสำเร็จรูปที่เป็นระบบฟรอนต์เอนด์เพื่อใช้ในการค้นคืนข้อมูลจากระบบฐานข้อมูลเชิงสัมพันธ์ที่รองรับการใช้ภาษาสอบถามแบบโครงสร้างหรือเอสคิวแอล (SQL - Structured Query Language) เช่น อินเกรส (Ingres), โอราเคิล (Oracle), ไซเบส (Sybase), อินฟอร์มิกส์ (Informix) เป็นต้น โดยมีความสามารถในการรับประโยคภาษาอังกฤษที่เป็นแบบภาษาฟอร์มัล (Formal Language) แล้วทำการแปลงประโยคให้เป็นภาษาเอสคิวแอล นอกจากนี้โปรแกรมยังสามารถเรียนรู้ศัพท์เฉพาะหรือรูปแบบการใช้ในประโยคได้ และให้ข้อมูลในรูปแบบรายงานต่างๆได้โดยไม่ต้องมีการพัฒนาโปรแกรมเพิ่มเติม โดยให้ผู้ใช้ออกคำสั่งทำรูปแบบรายงานตามความต้องการ สำหรับระบบฟรอนต์เอนด์ดังกล่าวที่รับคำสั่งเป็นประโยคภาษาไทยยังไม่มีหน่วยงานหรือองค์กรใดจัดทำขึ้นมา

ตัวอย่างงานวิจัยที่ใช้ภาษาธรรมชาติมาทำการค้นคืนข้อมูลในระบบฐานข้อมูล เช่น Cercone N. (1985) ได้กล่าวถึงความเป็นไปได้ในการจัดทำระบบค้นคืนข้อมูลด้วยภาษา

ธรรมชาติ และนำเสนอแนวทางการออกแบบระบบการตอบรับที่เป็นภาษาธรรมชาติเช่นกัน ส่วนงานของ Clifford J. (1988) ได้แสดงถึงวิธีการนำส่วนอรรถศาสตร์ของฐานข้อมูลและของภาษาธรรมชาติที่ใช้ในการขอค้นคืนมาเชื่อมเข้าด้วยกันโดยใช้ทฤษฎีอรรถศาสตร์ Montague โดยนำไปใช้กับระบบค้นคืนข้อมูลที่ชื่อ QE-III ซึ่งเป็นระบบที่ใช้ภาษาอังกฤษแบบฟอร์มัลในการค้นคืนข้อมูล และงานของ Davidson J. (1983) ที่นำเสนอการระบบต้นแบบที่ชื่อ PIQUE ใช้ในงานค้นคืนข้อมูลและการแก้ไขข้อมูลของฐานข้อมูล ส่วนงานวิจัยที่ใช้ทฤษฎีเอชพีเอสจีมาประยุกต์กับระบบค้นคืนข้อมูล เช่น งานของ McFetridge P. (1985) ที่นำเสนอผลงานการใช้เอชพีเอสจีมาใช้ในงานค้นคืนข้อมูลด้วยภาษาอังกฤษ

สำหรับงานวิจัยในประเทศไทยที่นำทฤษฎีเอชพีเอสจีมาใช้เช่นงานของ Wuwongse V. (1991) โดยนำเอชพีเอสจีมาใช้แก้ปัญหาการแบ่งคำ โดยนำระบบที่ชื่อ PATR มาใช้ในงานวิจัย นอกนั้นไม่พบว่ามีงานชิ้นใดที่นำเอาทฤษฎีเอชพีเอสจีมาใช้ในงานด้านการค้นคืนข้อมูลมาก่อนแต่อย่างใด

วัตถุประสงค์

ในงานวิจัยนี้จะทำการศึกษาความเป็นไปได้ในการนำทฤษฎีเอชพีเอสจี (HPSG - Head-driven Phrase Structure Grammar) มาใช้เป็นเครื่องมือในการวิเคราะห์กระจายประโยค และศึกษาหาแนวทางด้านเทคนิคที่จะนำมาประยุกต์ใช้ โดยนำรูปแบบไวยากรณ์และข้อมูลทางภาษาศาสตร์ของคำในประโยคภาษาไทยที่กำหนดโดยนักภาษาศาสตร์มาใช้งานในการประมวลผลในระบบฟรอนต์เอนด์เพื่อค้นคืนข้อมูล และศึกษาวิธีการตีความการอ้างถึงคำในประโยคแวดล้อมของส่วนแปลความหมายพอสังเขป

ขอบเขตของการวิจัย

1. อุปกรณ์ที่ใช้ในงานวิจัยนี้คือเครื่องคอมพิวเตอร์ระดับพีซี ที่มีไมโครโพรเซสเซอร์ 486 หรือดีกว่า

2. งานวิจัยครั้งนี้เน้นในการศึกษาส่วนวิเคราะห์ประโยคภาษาไทยโดยนำทฤษฎีเอชพีเอสจีมาใช้เป็นเครื่องมือในการวิเคราะห์กระจายประโยค เพื่อให้ได้ข้อมูลโครงสร้างองค์ประกอบของประโยค (Constituent Structure) ของประโยคในการนำมาแปลงเป็นคำสั่งสอบถามเพื่อค้นคืนข้อมูล

3. ศึกษาและทดลองหาเทคนิคการวิเคราะห์ประโยคภาษาไทยที่เหมาะสมกับเอชพีเอสจี

4. ประโยคภาษาไทยที่นำใช้เป็นคำสั่งตัวอย่างในการค้นคืนข้อมูลอยู่ในรูปแบบภาษาฟอร์มัล และนำมาใช้กับงานฐานข้อมูลประเภทใดประเภทหนึ่งเท่านั้น

5. ภาษาคำสั่งสอบถามที่นำมาใช้อ้างอิงในการสร้างคำสั่งสอบถามคือเอสคิวแอล ซึ่งเป็นภาษาสอบถามมาตรฐานที่ใช้กับระบบฐานข้อมูลเชิงสัมพันธ์ (Relational Database System) โดยทั่วไป

6. ในงานวิจัยครั้งนี้จะไม่ค้นคว้าหาวิธีการแบ่งคำในประโยคภาษาไทยแต่จะจำลองส่วนแบ่งคำไว้พอสังเขป โดยนำโปรแกรมการแบ่งคำไทยด้วยพจนานุกรมของนายสัมพันธ์ ระรื่นรมย์ (2533) มาเป็นตัวอย่างการใช้งานวิจัยนี้

ขั้นตอนของการวิจัย

1. ศึกษาทฤษฎีเอชพีเอสจี และค้นคว้าหาข้อมูลตัวอย่าง โดยทำการศึกษาทฤษฎีเอชพีเอสจีให้ละเอียดมากยิ่งขึ้น และพยายามหาตัวอย่างประโยคสอบถามในการใช้ค้นคืนข้อมูลจากระบบฐานข้อมูลให้มากที่สุด

2. สร้างส่วนแบ่งคำ โดยทำการปรับปรุงเพิ่มเติมจากงานของสัมพันธ์ ระรื่นรมย์ให้สามารถแบ่งคำให้ได้ถูกต้องยิ่งขึ้น

3. สร้างส่วนวิเคราะห์ประโยค โดยนำรูปแบบประโยคและข้อมูลทางภาษาศาสตร์ที่นักภาษาศาสตร์กำหนดไว้มาใช้ และปรับปรุงหรือเพิ่มเติมเท่าที่จำเป็น เพื่อให้ระบบสามารถทำการวิเคราะห์ประโยคได้อย่างถูกต้อง

4. สร้างส่วนแปลความหมาย ส่วนแปลความหมายนี้ทำเป็นโปรแกรมที่มีฐานข้อมูลอย่างง่ายเพื่อนำมาทำเป็นตัวอย่างในการสร้างส่วนแปลความหมาย

5. สร้างส่วนสร้างคำสั่งสอบถาม โดยทำการเขียนโปรแกรมตัวอย่างให้สามารถแปลงเป็นคำสั่งสอบถามเอสคิวแอล ทั้งนี้ภาษาเอสคิวแอลที่สร้างจะอิงกับภาษาเอสคิวแอลที่เป็นมาตรฐานเท่านั้น และในการแปลงส่งนี้จะไม่มีการทำส่วนของการเพิ่มประสิทธิภาพให้กับรหัส (Code Optimization)

6. ทดสอบ ทำการทดสอบการทำงานโดยทดลองรวมส่วนต่างๆให้ทำงานเป็นระบบฟรอต์เอนด์

ประโยชน์ที่ได้จากการวิจัย

ประโยชน์ที่คาดว่าจะได้รับจากงานวิทยานิพนธ์ชิ้นนี้ คือ

1. เป็นแนวทางในการนำเอาความรู้ทางด้านเอ็นแอลพีมาใช้งานในด้านตัวประสานกับผู้ใช้ในระบบฐานข้อมูล

2. สามารถนำงานวิจัยนี้ไปประยุกต์ใช้กับระบบฟรอนต์เอนด์ของระบบฐานข้อมูลจริงได้ โดยผู้ใช้ที่ไม่ชำนาญการใช้เครื่องคอมพิวเตอร์มากนักสามารถใช้งานด้านฐานข้อมูลได้โดยง่าย และปราศจากความกังวลในการทำงาน