



บทที่ 1

บทนำ

1.1 ความเบื้องต้น

การพัฒนาด้านคอมพิวเตอร์ได้มีการพัฒนาทางฮาร์ดแวร์ และซอฟต์แวร์ควบคู่กันไป จึงได้เกิดมีภาษาคอมพิวเตอร์ต่างๆ โดยเริ่มตั้งแต่ในยุคแรก ที่ใช้ภาษาเครื่องเป็นเลขฐานสอง พัฒนาขึ้นเป็นภาษาแอสเซมบลี และเป็นภาษาคอมพิวเตอร์ต่างๆ เช่น FORTRAN, COBOL หรือ ภาษาใหม่ๆ เช่น BASIC, Pascal, C, Lisp, Prolog, Ada ฯลฯ ซึ่งเป็นภาษาที่ใช้กันอยู่ ซึ่งภาษาเหล่านี้สามารถใช้งานทั่วไป โดยมีจุดเด่นที่ต่างกันไป ได้มีการเรียกภาษาเหล่านี้เป็นภาษารุ่นที่ 3 (Third Generation Language, 3rd GL)

ในขณะที่มีการใช้คอมพิวเตอร์กว้างขวางขึ้น งานที่ต้องการใช้คอมพิวเตอร์เป็นอย่างมากคือ งานทางด้านการเก็บรวบรวมข้อมูล ในขณะที่มีข้อมูลเอกสารเป็นจำนวนมาก ถ้าใช้คนจัดเก็บก็จะยุ่งยากเสียเวลามาก จึงได้มีการพัฒนาคอมพิวเตอร์เข้ามาช่วยจัดการกับข้อมูลเหล่านี้ ในการจัดการกับข้อมูลต่างๆ เราอาจเขียนโปรแกรมขึ้นมาเองจากภาษาต่างๆ เหล่านี้ โดยรวมข้อมูลแต่ละกลุ่มเป็นแฟ้มข้อมูลหนึ่ง และมีโปรแกรมจัดการเฉพาะสำหรับข้อมูลนั้น ถ้ามีการเพิ่มเติมข้อมูลชนิดอื่น ก็เขียนโปรแกรมขึ้นมาจัดการกับข้อมูลนั้นอีก ด้วยวิธีนี้ก็สามารถให้คอมพิวเตอร์ช่วยในการจัดการกับข้อมูลได้ แต่จะเห็นว่ามีย่อเสียอยู่คือ ต้องมีการเขียนโปรแกรมขึ้นมาจัดการเฉพาะสำหรับข้อมูลหนึ่งๆ ทำให้เป็นการสิ้นเปลืองเวลาในการพัฒนาโปรแกรม และยังมี ความซ้ำซ้อนของข้อมูล ด้วยเหตุนี้จึงได้มีการพัฒนาเป็นระบบฐานข้อมูลขึ้นมา เพื่อช่วยในการจัดการข้อมูลต่างๆ ให้เป็นระบบเดียวกัน

ระบบฐานข้อมูลนั้นเป็นการให้คอมพิวเตอร์ช่วยในการจัดเก็บระเบียบเท่านั้น (computerize record keeping) ซึ่งมีการทำงานที่ต้องการคือ การเพิ่มแฟ้มข้อมูลเข้าไปในฐานข้อมูลนั้น การเพิ่มระเบียบเข้าไปที่แฟ้มข้อมูลที่มีอยู่ การเลือกดึงระเบียบจากแฟ้มข้อมูลที่มี การแก้ไขระเบียบในแฟ้มข้อมูลนั้น การลบระเบียบที่ไม่ต้องการออก และการลบแฟ้มข้อมูลนั้นออกจากฐานข้อมูล [1]

การรวมข้อมูลเป็นอันหนึ่งอันเดียวกันนี้ทำให้ระบบฐานข้อมูลที่ทำให้คอมพิวเตอร์ช่วยทำ
ดีกว่าเมื่อเปรียบเทียบกับระบบที่ทำงานเอกสารคือ

- ความกะทัดรัด (compactness) ข้อมูลที่เก็บภายในคอมพิวเตอร์จะรวบรวมไว้
เป็นระเบียบ ทำให้ประหยัดเนื้อที่ในการเก็บมากกว่าการเก็บเป็นเอกสารจำนวนมาก
 - ความเร็ว (speed) สามารถจัดเก็บ และค้นหาข้อมูล รวมทั้งสรุปได้เร็วกว่า
 - ลดการดูแลข้อมูล (less drudgery) ไม่ต้องเก็บเอกสารจำนวนมาก
 - ข้อมูลทันสมัยกว่า (currency) เนื่องจากความเร็วในการจัดเก็บ
- จากการรวมข้อมูลเป็นกลุ่มเดียวกัน ทำให้เป็นการควบคุมข้อมูลในการปฏิบัติการ
จากส่วนกลางได้

การใช้ระบบฐานข้อมูลนี้จะมีข้อดีมากกว่าการแยกกันจัดการข้อมูลแต่ละส่วนด้วย
โปรแกรมจัดการที่แตกต่างกันคือ

- ลดการซ้ำซ้อนของข้อมูล (redundancy)
- ลดความไม่เกี่ยวเนื่องของข้อมูล (inconsistency) เนื่องจากการซ้ำซ้อน
ของข้อมูลอาจทำให้ข้อมูลที่เกี่ยวเนื่องกันมีค่าที่แตกต่างกัน
- สามารถใช้ข้อมูลรวมกัน (shared)
- มีความเป็นมาตรฐานในการจัดเก็บ (standard) มีประโยชน์ในการแลกเปลี่ยน
ข้อมูลกัน
- จัดการด้านรักษาความปลอดภัยได้ (security)
- ข้อมูลมีความถูกต้อง (integrity) ทำให้ข้อมูลเชื่อถือได้ (accurate)
- ลดปัญหาความขัดแย้ง (conflict balance) เนื่องจากสามารถรองรับงานทั้ง

ระบบ

ระบบฐานข้อมูลมีได้หลายแบบ ตามแต่การจัดเก็บ และความเกี่ยวเนื่องของข้อมูล
ภายใน เช่นการจัดข้อมูลแบบสัมพันธ์ (relational model) การจัดข้อมูลแบบโครงข่าย
(network model) และการจัดข้อมูลแบบเป็นลำดับชั้น (hierarchical model) เป็นต้น
แต่ที่จะกล่าวถึงต่อไปจะกล่าวถึงระบบฐานข้อมูลแบบสัมพันธ์

แนวความคิดของระบบฐานข้อมูลแบบสัมพันธ์นี้ได้รับการเสนอจาก Dr. E.F Codd
[2] ในขณะที่เป็นสมาชิกของศูนย์วิจัยของบริษัท IBM ที่ San Jose, California ระหว่าง
ค.ศ. 1969-1970 โดยเสนอความคิดทางคณิตศาสตร์ในการจัดการระบบข้อมูล ความคิดนี้

เป็นที่ยอมรับในวงการศึกษาทั่วไป และได้มีการวิจัยเพื่อสร้างระบบต้นแบบขึ้นในช่วงกลางทศวรรษ 1970 ระบบที่พัฒนาขึ้นนี้มีมากมายแต่จะขอกล่าวถึงประวัติความเป็นมาของระบบฐานข้อมูลใช้กันแพร่หลาย 2 ระบบ คือ

ก. System R ระบบที่พัฒนาขึ้นที่ศูนย์วิจัยของ IBM ที่ San Jose ในระหว่าง ค.ศ. 1974-1977

ข. University Ingres พัฒนาค้นที่ University of California, Berkeley ระหว่าง ค.ศ. 1973-1975

ระยะเวลาที่กำหนดขึ้นเป็นระยะเวลาโดยประมาณ ทั้งสองระบบก็ยังมีพัฒนาต่อเนื่องมาจนถึงปัจจุบัน ทั้งสองระบบนี้เป็นการวิจัยเพื่อสร้างต้นแบบระบบฐานข้อมูลที่มีการทำงานครบถ้วน เช่น มีการทำ locking, journaling, recovery ฯลฯ โดยคำนึงถึงประสิทธิภาพสูงสุด แต่ยังไม่ใช้ได้ง่าย

System R ได้พัฒนาต่อมาเป็น DB2 (และ SQL/DS) [3] โดย IBM และ INGRES พัฒนาคือ Commercial INGRES [4] โดยบริษัท RTI ได้มีการพัฒนาภาษาการสอบถามขึ้นใช้ในการจัดการกับระบบฐานข้อมูลขึ้น คือ System R ได้พัฒนาภาษา SQL (Structure Query Language) และ INGRES ได้พัฒนาภาษา QUEL (Query Language) ซึ่งต่อมาได้มีการกำหนดมาตรฐานทางภาษาการสอบถาม โดย ANSI ได้กำหนดจากภาษา SQL ไปใช้ในปี ค.ศ. 1986

นอกจากทั้งสองระบบที่กล่าวมาแล้วยังมีระบบอื่นอีกมาก ทั้งบนเครื่องคอมพิวเตอร์ระดับเมนเฟรม จนถึงไมโครคอมพิวเตอร์ ตัวอย่างของระบบฐานข้อมูลแบบสัมพันธ์เช่น System R, SQL/DS, DB2, Oracle, Ingres ฯลฯ ระบบฐานข้อมูลแบบโครงสร้างเช่น TOTAL, IDMS ฯลฯ ระบบฐานข้อมูลแบบลำดับขั้นเช่น IMS, System 2000 ระบบบนเครื่องไมโครคอมพิวเตอร์เช่น dBASE, PARADOX, PFS เป็นต้น

ระบบฐานข้อมูลแบบสัมพันธ์ จะเป็นความสัมพันธ์ภายในของข้อมูล ในการพิจารณา ระบบฐานข้อมูลแบบสัมพันธ์เราอาจดูได้จาก ข้อมูลมีการเก็บในรูปของตาราง ผู้ใช้มองไม่เห็นดัชนี หรือ ตัวเชื่อมโยงภายใน มีภาษาการสอบถาม และในการสอบถาม ไม่จำเป็นต้องใช้วิธีการ iteration [5]

Codd ได้กำหนดความหมายของคำขึ้นมาเพื่อให้มีความแตกต่างจากระบบเดิม เช่น
- relation หมายถึง table

- tuple หมายถึง record หรือ row
- attribute หมายถึง field หรือ column
- primary key หมายถึง unique identifier

ภาษาการสอบถาม จะช่วยให้ผู้ใช้สามารถจัดการกับข้อมูลภายในฐานข้อมูลได้ง่าย การทำงานภายในจะเป็นในรูปของเซต จัดการกับข้อมูลได้เป็นกลุ่มซึ่งแตกต่างจากภาษาตั้งเดิม มีการจัดให้เป็นภาษารุ่นที่ 4 (4th GL) ซึ่งใช้ทางด้านระบบฐานข้อมูล นอกจากนี้ยังมีภาษารุ่นที่ 5 (5th GL) ซึ่งพัฒนาในด้านปัญญาประดิษฐ์ (Artificial Intelligent, AI) และได้มีการวิจัยที่จะรวมด้านปัญญาประดิษฐ์ช่วยจัดการระบบฐานข้อมูล [6]

ทางด้านความคิดต่อกับผู้ใช้ จะเห็นว่าภาษาต่างๆ ที่ใช้ก็เพื่อช่วยในการติดต่อระหว่างผู้ใช้กับระบบนั้น ได้มีการพัฒนาคอมพิวเตอร์เพื่อช่วยในการจัดการกับเอกสารต่างๆ (Office Automation, OA) จึงได้เกิดความคิดในการใช้ฟอร์มเพื่อแทนเอกสารเหล่านั้น [7] โดยใช้ฟอร์มเป็นสื่อในการติดต่อกับผู้ใช้อย่างหนึ่ง โดยไม่ขึ้นกับชนิดของสื่อ นั้น เช่น ทางภาพ หรือ เสียง เป็นต้น ภายในฟอร์มจะประกอบด้วย ส่วนคงที่ที่ปรากฏในทุกฟอร์ม (template) และ ส่วนช่องว่าง (slot) ที่เป็นส่วนติดต่อกับข้อมูลจากระบบฐานข้อมูล การใช้ฟอร์มในรูปแบบต่างๆ ก็เป็นที่ยอมรับกันทั่วไป

ในการพัฒนาโปรแกรมขึ้นมา นั้น ได้แบ่งการพัฒนาโปรแกรมออกเป็น 2 ส่วน คือ ส่วนที่เป็นโครงสร้างข้อมูล (data structure) และการขั้นตอนการทำงาน (algorithm) [8] การพัฒนาโปรแกรมส่วนใหญ่ก็มีรูปแบบเป็นเช่นนี้ ได้มีการพัฒนาแนวความคิดในการมองข้อมูลในรูปแบบของการกำหนดชนิดข้อมูลแบบย่อ (abstract data type) โดยได้มีการให้คำจำกัดความไว้ว่า abstract data type ในกลุ่มของผล และการกระทำ ซึ่งการใช้งานนั้นจะซ่อนไว้จากผู้ใช้งาน ผู้ใช้จะเห็นเพียงแต่ผล และการกระทำที่ได้จากการใช้งานในการกระทำต่อผลนั้นเท่าที่กำหนดให้เท่านั้น [9] การกำหนดชนิดข้อมูลแบบย่อช่วยในการพัฒนาโปรแกรมให้เร็วขึ้น เนื่องจากสามารถกำหนดชนิดของข้อมูลในภายหลัง เช่น ในขณะคอมไพล์ (compile) หรือใช้งาน (run time) แต่อาจทำงานได้ช้าลงบ้าง

ในประเทศไทยได้มีการใช้งานทางระบบฐานข้อมูลมานานแล้ว แต่ยังคงขาดการวิจัยและพัฒนาฐานข้อมูลภาษาไทยอยู่มาก ภาษาไทยมีเอกลักษณ์ที่แตกต่างจากภาษาอื่นๆ เช่น ภาษาอังกฤษ ทำให้ต้องมีการกำหนดรหัสเพื่อใช้แทนตัวอักษร การจัดเรียงลำดับอักษรเพื่อแสดงผล การกำหนดตำแหน่งแบ่นพิมพ์ และพัฒนาอัลกอริทึมการเปรียบเทียบคำไทยเพื่อใช้

ในการเรียงลำดับคำ [10] ได้มีการกำหนดรหัสอักขระเป็นมาตรฐานในการแลกเปลี่ยนข้อมูลภาษาไทยใน พ.ศ. 2529 [11] แต่ยังคงขาดการกำหนดมาตรฐานรหัสการแสดงผล และตำแหน่งแป้นพิมพ์ เนื่องจากขึ้นอยู่กับอุปกรณ์ที่คอมพิวเตอร์นั้นต่ออยู่ ในภาษาไทยมีการจัดเรียงตำแหน่งสระบน สระล่าง ต่างจากการเรียงแถวเดียวในภาษาอังกฤษ ทำให้ต้องมีการดัดแปลงอุปกรณ์ต่างๆ ที่ต้องการให้แสดงภาษาไทย หรือปรับปรุงโปรแกรมให้แตกต่างไปจากเดิม

ต่อไปจะกล่าวถึงการจัดการระบบฐานข้อมูลเฉพาะบนระบบ ไมโครคอมพิวเตอร์ ในการพัฒนาเพื่อให้ใช้ภาษาไทยกับโปรแกรมสำเร็จรูปต่างๆ นั้น ได้มีการดัดแปลงทั้งฮาร์ดแวร์ของวงจรมาคแสดง หรือ ดัดแปลงตัวโปรแกรมเองด้วย การดัดแปลงนี้มีผู้ทำกันมากทำให้แต่ละระบบที่ถูกดัดแปลงแล้วไม่สามารถใช้แทนกันได้ และยังอาจมีปัญหานในข้อจำกัดของการดัดแปลงเนื่องจากการปกปิดข้อมูลของฮาร์ดแวร์ หรือซอฟต์แวร์นั้นๆ การดัดแปลงจึงอาจทำให้เกิดผลข้างเคียงอื่นได้ ดังนั้นจึงควรวิจัยและพัฒนาโปรแกรมสำหรับภาษาไทยโดยเฉพาะขึ้นใช้เอง

ในทางด้าน ไมโครคอมพิวเตอร์มีข้อจำกัดต่างๆ มากกว่าเครื่องใหญ่ เช่นขนาดของข้อมูล และความซับซ้อนของระบบ ผู้ใช้เป็นผู้ควบคุมเครื่องเองทั้งหมดแต่ผู้เดียว ผู้ใช้จึงไม่จำเป็นต้องมีความเชี่ยวชาญเฉพาะ ดังนั้นระบบฐานข้อมูลบน ไมโครคอมพิวเตอร์จึงแตกต่างจากระบบฐานข้อมูลระบบใหญ่ ซึ่งพอจะเปรียบเทียบได้ดังนี้

- การกำหนดรูปแบบข้อมูล ระบบฐานข้อมูล ไมโครคอมพิวเตอร์เป็นระบบที่ค่อนข้างใหม่ จึงมีการจัดรูปแบบข้อมูลเป็นชนิดสัมพันธ์
- ภาษาการสอบถาม อาจซับซ้อนเกินไป จึงใช้ในรูปแบบการติดต่อกับผู้ใช้ผ่านทางฟอร์ม และเป็นการเติมข้อมูลในฟอร์มนั้น
- การจัดการข้อมูลจริง ถ้ามีการใช้ตัวจัดการหน่วยความจำสำรอง การใช้ดัชนีจำนวนมาก ขนาดโปรแกรมจะมีขนาดใหญ่ กินเนื้อที่หน่วยความจำ
- การดึงข้อมูลกลับเมื่อข้อมูลเสีย ไม่มีความจำเป็น สามารถใช้การสำรองข้อมูลทั้งระบบได้
- ความทันสมัยของข้อมูล ไม่ต้องคำนึงถึงเนื่องจากมีผู้ใช้เพียงคนเดียว

จากปัญหาดังกล่าว ดังนั้นจึงได้เกิดการพัฒนารวมการเพิ่มข้อมูลขึ้นบน ไมโครคอมพิวเตอร์ เพื่อนำไปใช้งานเป็นเครื่องมือทางฐานข้อมูล และเป็นพื้นฐานในการพัฒนาต่อไป

1.2 วัตถุประสงค์ของการวิจัย

ทำการออกแบบ และพัฒนา เครื่องมือทางซอฟต์แวร์ เพื่อช่วยในการทำงานกับระบบฐานข้อมูลไทย/อังกฤษ ซึ่งเครื่องมือนี้จะช่วยทำงานหลักๆ เช่น สร้างเพิ่มข้อมูล, เพิ่มเติมข้อมูล, ลบข้อมูล, ค้นหาและแก้ไขข้อมูล โดยใช้ได้กับเพิ่มข้อมูลไม่จำกัดจำนวน และสามารถกำหนดโครงสร้างของเพิ่มข้อมูล (File structure) ได้อย่างอิสระ

ขอบเขตการทำงาน และ ข้อกำหนดของโปรแกรม

1. เลือกเครื่องมือในการพัฒนาโปรแกรมที่จำเป็น เช่นใช้ภาษา Pascal หรือ C
2. ทำงานบนระบบเครื่องไมโครคอมพิวเตอร์ IBM PC (XT/AT or Compatible) โดยไม่ต้องมีการดัดแปลง หรือเพิ่มเติมทางฮาร์ดแวร์
3. การจัดการเพิ่มข้อมูลกำหนดผ่านทางจอภาพ โดยผู้ใช้จัดวางข้อความบนจอภาพเพื่อให้สัมพันธ์กับเพิ่มข้อมูลที่ต้องการ โดยผู้ใช้ไม่จำเป็นต้องรู้วิธีเก็บข้อมูลจริง
4. ผู้ใช้ต้องกำหนด ชื่อเพิ่มข้อมูล, โครงสร้างของข้อมูล และ การจัดวางข้อความบนจอภาพ
5. การจัดการเกี่ยวกับจอภาพเป็น Full screen edit
6. สามารถจัดการกับข้อมูลไทย/อังกฤษ และตรวจสอบความเป็นไปได้ของค่าไทย
7. สามารถ สร้างเพิ่มข้อมูล, เพิ่มเติมข้อมูล, ลบข้อมูล, ค้นหา และแก้ไขข้อมูลที่ต้องการได้ โดยไม่จำกัดจำนวนเพิ่มข้อมูล แต่สามารถทำงานได้ทีละเพิ่มข้อมูลเดียว

1.3 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1. เพื่อนำเครื่องมือที่ได้ไปประยุกต์ใช้งาน
2. เป็นพื้นฐานในการพัฒนางานทางด้าน ระบบจัดการฐานข้อมูลต่อไป เช่น พัฒนาโปรแกรมผู้สร้างรายงาน (report generator) ที่ใช้แนวความคิดเดียวกันได้ ตลอดจนขยายไปถึงระบบจัดการฐานข้อมูลภาษาไทย
3. เป็นการพัฒนาเครื่องมือทางซอฟต์แวร์ (software tool) สำหรับงานอื่นต่อไป เช่น ส่วนจัดการเกี่ยวกับระบบภาษาไทย