

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการพัฒนาระบบป้องกันสแปมเมลล์สำหรับจดหมายอิเล็กทรอนิกส์ของบุคลากรและนักศึกษามหาวิทยาลัยเชียงใหม่ ผู้ศึกษาได้ค้นคว้าเอกสารและงานวิจัยที่เกี่ยวข้อง โดยมีรายละเอียดตามหัวข้อที่กำหนดตามลำดับดังนี้

2.1 แนวความคิดเกี่ยวกับระบบฐานข้อมูล

2.1.1 ความหมายของระบบฐานข้อมูล

ศิวัช กาญจนชุม และวิชาญ หงษ์บิน (2542) ได้ให้ความหมายของฐานข้อมูล (Database) หมายถึงกลุ่มของข้อมูล (Data Group) ที่ถูกรวบรวมเข้าไว้ด้วยกัน โดยครอบคลุมรายละเอียดต่าง ๆ เช่น ในสำนักงานก็จะรวบรวมข้อมูลตั้งแต่หมายเลขโทรศัพท์ของผู้ที่มาติดต่อจนถึงการเก็บเอกสารทุกอย่างของสำนักงาน ซึ่งข้อมูลจะมีส่วนที่สัมพันธ์กันและเป็นที่ต้องการออกมาใช้ให้เป็นประโยชน์ ข้อมูลนั้นอาจจะเกี่ยวกับบุคคล สิ่งของ สถานที่ หรือเหตุการณ์ใด ๆ ที่เราสนใจศึกษา ซึ่งข้อมูล (Data) อาจจะได้มาจากการสังเกต การนับหรือการวัด และข้อมูลอาจเป็นได้ทั้งตัวเลขหรือเป็นข้อความก็ได้ ที่สำคัญคือข้อมูลจะต้องเป็นสิ่งที่ เป็นความจริง

2.1.2 องค์ประกอบของระบบฐานข้อมูล

องค์ประกอบของระบบฐานข้อมูลในที่นี้ หมายถึง โครงสร้างสารสนเทศ (Information) ที่ประกอบด้วยเอนทิตี (Entity) หลาย ๆ ตัว ซึ่งบรรดาเอนทิตีเหล่านี้จะต้องมีความสัมพันธ์กัน

เอนทิตีในที่นี้หมายถึง สิ่งที่มีอยู่จริง อันได้แก่ บุคคล สถานที่ และสิ่งของ

ตัวอย่างเอนทิตีในระบบฐานข้อมูลของโรงเรียนอาจได้แก่ อาจารย์ แผนกวิชา ประวัติการทำงาน หรือถ้าเป็นเอนทิตีของบริษัทก็อาจจะได้แก่ พนักงานขาย ลูกค้า การสั่งซื้อ และประเภทสินค้า เป็นต้น

ส่วนของข้อมูลที่แสดงลักษณะและคุณสมบัติของเอนทิตีเรียกว่า แอททริบิวต์ (Attribute) เช่น

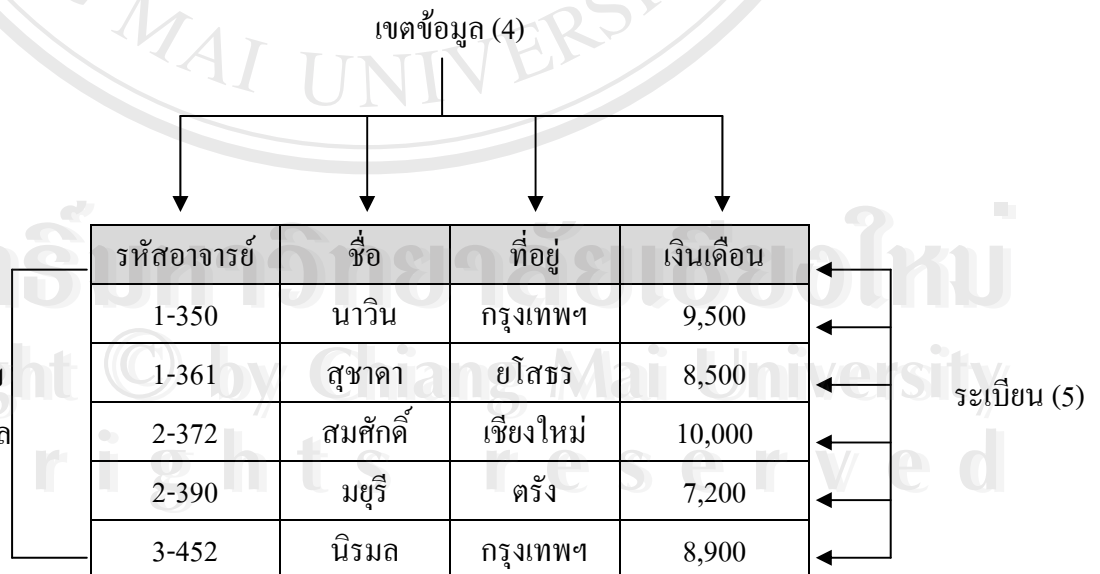
- แอททริบิวต์ของบุคคลได้แก่ ชื่อ เพศ อายุ ฯลฯ
- แอททริบิวต์ของอาจารย์ได้แก่ รหัสอาจารย์ ชื่อ ที่อยู่ เงินเดือน ฯลฯ
- แอททริบิวต์ของประเภทสินค้า ได้แก่ รหัสสินค้า สี ราคา ฯลฯ

ดังนั้น ถ้าจะเปรียบเทียบไปแล้วจะเห็นว่าเอนทิตีนั้นเหมือนกับแฟ้มข้อมูล ส่วนแอททริบิวต์นั้นเหมือนกับเขตข้อมูลนั่นเอง

เขตข้อมูล หรือ ฟิลด์ (Field) หมายถึง ตัวอักขระ (Character) แต่ละตัว ถูกนำมาประกอบกันเป็นกลุ่มคำที่มีความหมายขึ้น เช่น กลุ่มตัวอักษรที่ประกอบกันเป็นชื่อหรือนามสกุล กลุ่มตัวเลขที่ประกอบกันเป็นรหัสประจำตัว, ราคา โดยเรียกกลุ่มข้อมูลนี้ว่า เขตข้อมูล เช่น เขตข้อมูลชื่อ เขตข้อมูลนามสกุล เขตข้อมูลรหัสประจำตัว เป็นต้น

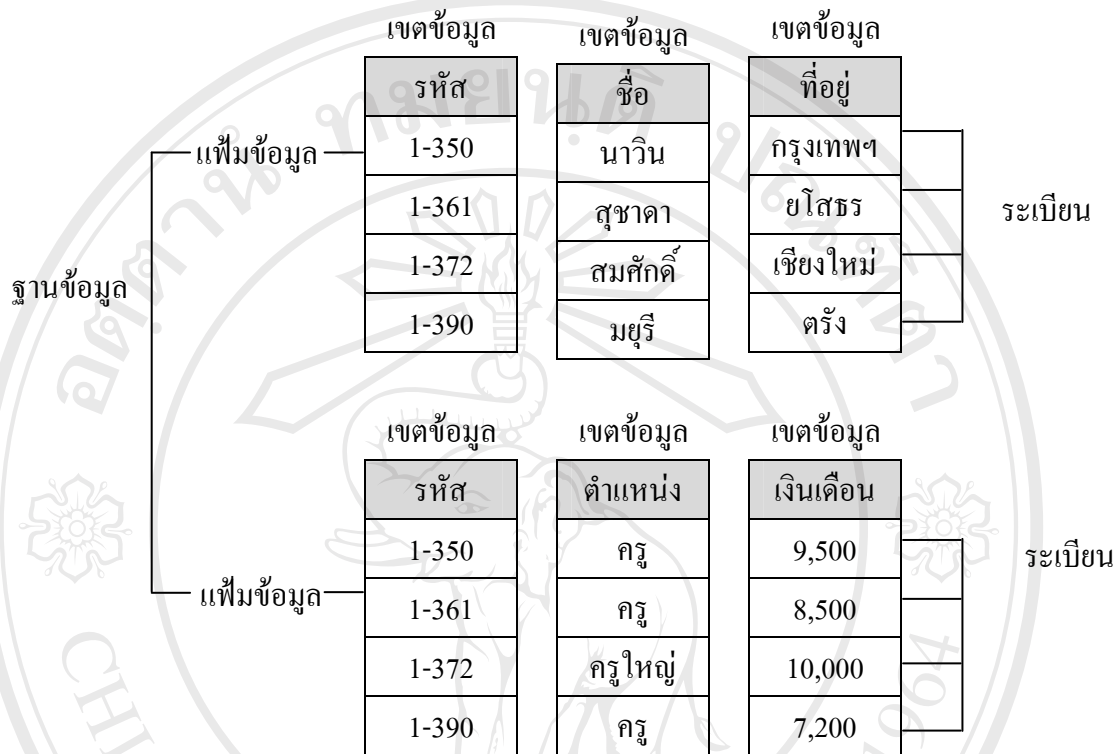
ตัวอักขระ คือข้อมูลพื้นฐานที่เล็กที่สุดภายในแฟ้มข้อมูลคือ บิต (Bit : Binary Digit) ซึ่งเป็นหน่วยข้อมูลพื้นฐานที่เก็บอยู่ในหน่วยความจำภายในคอมพิวเตอร์ บิตนี้จะแทนด้วยตัวเลข 1 ตัวได้แก่ 1 หรือ 0 ใดๆอย่างหนึ่ง เรียกตัวเลข 1 หรือ 0 นี้ว่า เป็นบิต 1 บิต ข้อมูลซึ่งได้แก่ ตัวอักขระแต่ละตัวเช่น A, B,Z, 0, 1, 2, 9 และสัญลักษณ์พิเศษอื่น ๆ เช่น \$, &, +, -, *, / ฯลฯ เมื่อจะถูกนำไปเก็บไว้ในคอมพิวเตอร์จะต้องถูกแปลงให้อยู่ในรูปของบิตหลายบิตที่มาประกอบกัน โดยตัวอักขระ 1 ตัวจะแทนด้วยบิต 7 หรือ 8 บิต ตัวอักขระแต่ละตัวจะเรียกได้อีกชื่อว่า ไบท์ (Byte) ตัวอย่างเช่น ตัวอักขระ A เมื่อเก็บอยู่ในคอมพิวเตอร์จะเก็บเป็น 1000001 ตัวอักขระ B จะเก็บเป็น 100010 เป็นต้น

เมื่อนำข้อมูลในหลายเขตข้อมูลหรือฟิลด์มารวมกันซึ่งมีลักษณะเป็นแถว ๆ จะเรียกว่า ระเบียบหรือเรคคอร์ด (Record) ระเบียบแต่ละระเบียบของข้อมูลชนิดเดียวกัน จะสามารถนำมา รวมกันเป็นแฟ้มข้อมูลหรือไฟล์ (File) ดังตัวอย่างต่อไปนี้



รูป 2.1 แสดงรายละเอียดของแฟ้มข้อมูล

ถ้า นำเพิ่มข้อมูลที่มีความสัมพันธ์กัน ซึ่งแต่ละเพิ่มข้อมูลจะประกอบด้วยเขตข้อมูลมา รวมกันนั้นเรียกว่า ฐานข้อมูล



รูป 2.2 แสดงความสัมพันธ์ของตาราง

สรุปได้ว่า ถ้า นำเพิ่มข้อมูลหลายเพิ่มข้อมูลมารวมกันก็จะกลายเป็นฐานข้อมูลได้ แต่ ฐานข้อมูลที่สมบูรณ์จะต้องมีการเก็บคำอธิบายเกี่ยวกับโครงสร้างของฐานข้อมูลที่เรียกว่า พจนานุกรมข้อมูล(Data Dictionary) หรืออาจเรียกอีกอย่างหนึ่งว่าเมตาดาตา (Meta Data) ด้วย

พจนานุกรมข้อมูล หมายถึง ส่วนที่มีหน้าที่อธิบายลักษณะของข้อมูลที่เก็บอยู่ใน ฐานข้อมูล รวมทั้งความสัมพันธ์ของข้อมูล

ระบบการจัดการฐานข้อมูล (Database Management System หรือ DBMS) ถ้ามีหลายตัว จะมีการรวมพจนานุกรมข้อมูลเป็นส่วนหนึ่งของ DBMS ด้วย ซึ่งพจนานุกรมข้อมูลนี้จะเป็น องค์ประกอบ ทางซอฟต์แวร์ ทำหน้าที่เก็บรายละเอียดเกี่ยวกับข้อมูลภายในฐานข้อมูล เช่น โครงสร้างของ แต่ละตาราง ใครเป็นผู้สร้าง สร้างเมื่อใด และแต่ละตารางประกอบด้วยเขต ข้อมูลใดบ้าง คุณลักษณะของแต่ละเขตข้อมูลเป็นอย่างไร มีการเรียกใช้อยู่ในโปรแกรมประยุกต์ ใดบ้าง และมีตารางใดที่มี ความสัมพันธ์กันบ้าง มีเขตข้อมูลใดเป็นคีย์บ้าง เป็นต้น

พจนานุกรมข้อมูลยังมีส่วนที่ทำหน้าที่สำคัญดังต่อไปนี้

- ความคุมระบบรักษาความปลอดภัยของข้อมูล และความคงสภาพของข้อมูล (Data Security and Data Integrity)
- ความคุมเกี่ยวกับการใช้งานฐานข้อมูลพร้อมกัน (Concurrency Control)

ถ้าเปรียบเทียบฐานข้อมูลเหมือนกับห้องสมุดของโรงเรียนที่มีการเก็บหนังสือเล่มต่าง ๆ หนังสือเหล่านั้นจะเปรียบเสมือนกับข้อมูลที่เก็บอยู่ภายในฐานข้อมูล และในห้องสมุดจะต้องมีการทำบัญชีรายชื่อหนังสือต่าง ๆ เก็บไว้ เพื่อใช้บอกรายละเอียดเกี่ยวกับหนังสือแต่ละเล่มว่าใครเป็นผู้แต่งเก็บอยู่ที่ใดในห้องสมุด บัญชีรายชื่อหนังสือนี้จะเปรียบได้กับพจนานุกรมข้อมูล เช่น ระหว่างระเบียบของแต่ละแฟ้มข้อมูล จะมีพจนานุกรมข้อมูลซึ่งถูกเก็บและจะถูกเรียกใช้งานในระหว่างที่มีการประมวลผลฐานข้อมูลและข้อมูลที่มีอยู่ในฐานข้อมูลจะต้องมีความสัมพันธ์กันด้วย

2.1.3 ประโยชน์จากการประมวลผลด้วยฐานข้อมูล

การประมวลผลด้วยระบบฐานข้อมูล มีประโยชน์ดังนี้

- 1) ลดความซ้ำซ้อนของข้อมูลในกรณีที่มีข้อมูลชนิดเดียวกันถูกเก็บไว้หลายๆแห่ง หรือที่เรียกว่าความซ้ำซ้อนการนำข้อมูลทั้งหมดมาเก็บไว้ที่เดียวกันในฐานข้อมูล จึงถือว่าเป็นการ “ลด” ความซ้ำซ้อนลงไปได้ทั้งนี้มิใช่หมายความว่าให้จัดข้อมูลออกไปเพื่อให้เหลือน้อยลง
- 2) สามารถหลีกเลี่ยงความขัดแย้งของข้อมูลได้ในระดับหนึ่งเนื่องจากบางครั้งจะต้องมีการแก้ไขข้อมูลจึงอาจจะก่อให้เกิดปัญหาในการแก้ไขคือเมื่อเราแก้ไขข้อมูลที่เหมือนกันแต่แก้ไขไม่หมดหรือแก้ไขไม่ครบทุกข้อมูลที่มีอยู่ในแต่ละแห่งจึงทำให้ข้อมูลชุดเดียวกันอาจมีค่าในแต่ละแห่งไม่ตรงกันดังนั้นถ้าการใช้ระบบฐานข้อมูลทำให้เราสามารถลดความซ้ำซ้อนลงไปได้ซึ่งถ้าใช้ระบบฐานข้อมูลเมื่อเกิดการแก้ไขข้อมูลขึ้นเมื่อใดก็จะต้องแก้ไขให้เหมือนกันครบทุกแห่ง
- 3) สามารถใช้ข้อมูลร่วมกันได้การใช้ข้อมูลร่วมกันได้นี้ไม่ได้จำกัดอยู่เฉพาะโปรแกรมที่ใช้ข้อมูลอยู่ในปัจจุบันเท่านั้นแต่โปรแกรมประยุกต์ที่มีการพัฒนาขึ้นมาใหม่ ก็สามารถใช้ข้อมูลที่มีอยู่ได้เลยโดยไม่จำเป็นต้องเพิ่มเติมข้อมูลเข้าไปในระบบอีก
- 4) สามารถควบคุมความเป็นมาตรฐานได้ เมื่อมีการนำข้อมูลมาเก็บรวบรวมกันไว้ในฐานข้อมูลเช่นนี้ทำให้ผู้ที่ทำหน้าที่ควบคุมดูแลการใช้ระบบฐานข้อมูลสามารถกำหนดมาตรฐานของข้อมูลขึ้นมาได้ทำให้การบริหารหรือแลกเปลี่ยนข้อมูลระหว่างระบบเป็นไปอย่างสะดวกและถูกต้อง
- 5) สามารถจัดหาระบบความปลอดภัยที่รัดกุมได้ระบบฐานข้อมูลสามารถที่จะกำหนดสิทธิการใช้ให้แก่ผู้ใช้คนใดก็ได้ตามความเหมาะสมและผู้ใช้แต่ละคนก็อาจจะใช้ข้อมูลได้

ในระดับที่ต่างกันหรือพูดอีกนัยหนึ่งก็คือผู้ใช้แต่ละคนจะมองฐานข้อมูลด้วยวิธีที่ต้องการจึงทำให้มีความปลอดภัยในการใช้ข้อมูลร่วมกัน

6) สามารถควบคุมความคงสภาพของข้อมูลได้โดยมีการกำหนดค่าต่างๆหรือจำกัดช่วงของข้อมูลไว้เพื่อป้องกันการพิมพ์ข้อมูลผิดพลาดเช่นกำหนดช่วงของข้อมูลในการกรอกหมายเลขโทรศัพท์ไว้ 7 ตัว เมื่อพิมพ์ครบ 7 ตัวแล้ว ก็กำหนดให้เลื่อนไปข้อมูลถัดไป ฉะนั้นถ้าพิมพ์ไม่ครบโปรแกรมก็จะไม่เลื่อนให้หรือพิมพ์เกิน 7 ตัวก็ไม่ได้

7) สามารถสร้างสมดุลในความขัดแย้งของความต้องการได้การใช้ที่ผู้ใช้ทั้งหมดขององค์กรใช้ข้อมูลจากฐานข้อมูลร่วมกันทำให้ทราบถึงความต้องการและความสำคัญของผู้ใช้งานทั้งหมดจึงสามารถกำหนดโครงสร้างของฐานข้อมูลเพื่อให้บริการที่ดีที่สุดได้เช่นเลือกเก็บข้อมูลที่จะต้องใช้บ่อยๆไว้ในสื่อข้อมูลที่มีความเร็วเป็นพิเศษเป็นต้นเป็นการสร้างสมดุลของความต้องการไม่ให้เกิดความขัดแย้งในหมู่ผู้ใช้เพราะการออกแบบนั้นกระทำบนแนวทางที่มุ่งจะให้ประโยชน์ส่วนรวม ดีที่สุดแล้ว

8) เกิดความเป็นอิสระของข้อมูลโปรแกรมประยุกต์ที่เขียนขึ้นจะไม่ขึ้นกับโครงสร้างของตารางที่มีการเปลี่ยนแปลงไปเนื่องจากโครงสร้างของตารางต่างๆ และตัวข้อมูลในแต่ละตารางจะถูกเก็บอยู่ในฐานข้อมูลทั้งหมดโปรแกรมประยุกต์ไม่จำเป็นต้องเก็บโครงสร้างของตารางที่จะใช้ไว้ ซึ่งต่างกับระบบการประมวลผลเพิ่มข้อมูล ดังนั้นการเปลี่ยนแปลงโครงสร้างของตารางเช่นการเปลี่ยนแปลงขนาดของเขตข้อมูลในตารางใดภายในฐานข้อมูล ก็ไม่จำเป็นต้องไปทำการแก้ไขโปรแกรมประยุกต์ที่มีการเรียกใช้เขตข้อมูลนั้น

2.1.4 ความสัมพันธ์และรูปแบบฐานข้อมูล

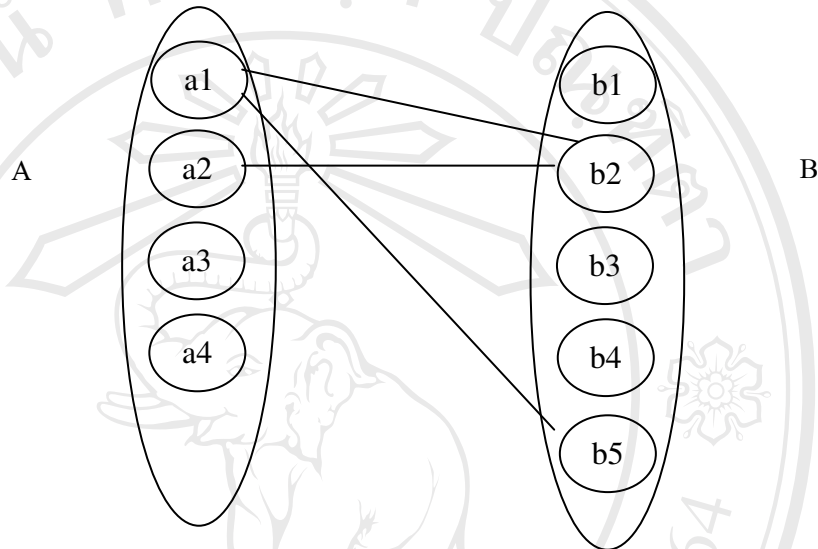
1) ความสัมพันธ์ (Relationship)

ฐานข้อมูลที่ใช้ในงานหนึ่ง ๆ นั้น ปกติมักจะประกอบด้วยแฟ้มข้อมูลหลายแฟ้ม แต่ละแฟ้มเก็บข้อมูลแตกต่างกันออกไป ฉะนั้นจึงต้องมีการกำหนดความสัมพันธ์ระหว่างข้อมูลต่าง ๆ ที่ต้องการเก็บ ซึ่งความสัมพันธ์นี้อาจจำแนกได้เป็น 3 ชนิด คือ

- ความสัมพันธ์แบบหนึ่งต่อหนึ่ง (One-to-One Relationship)
- ความสัมพันธ์แบบหนึ่งต่อกลุ่ม (One-to-Many Relationship)
- ความสัมพันธ์แบบกลุ่มต่อกลุ่ม (Many-to-Many Relationship)

ความสัมพันธ์แบบหนึ่งต่อหนึ่ง

เป็นลักษณะความสัมพันธ์ที่มีระเบียนเพียง 1 ระเบียนในเอนทิตี A ที่มีความสัมพันธ์กับระเบียนเพียง 1 ระเบียนในเอนทิตี B และในทางกลับกัน ระเบียนเพียง 1 ระเบียนในเอนทิตี B จะมีความสัมพันธ์กับระเบียนเพียง 1 ระเบียนในเอนทิตี A



รูป 2.3 แสดงความสัมพันธ์ในลักษณะหนึ่งต่อหนึ่ง

สมาชิก

การยืมหนังสือ

รหัสสมาชิก	ชื่อสมาชิก	แผนก
1540	อรนุช	การขาย
1645	กึ่งแก้ว	เลขานุการ
1748	ภาวนา	การบัญชี
1985	สมนึก	คอมพิวเตอร์

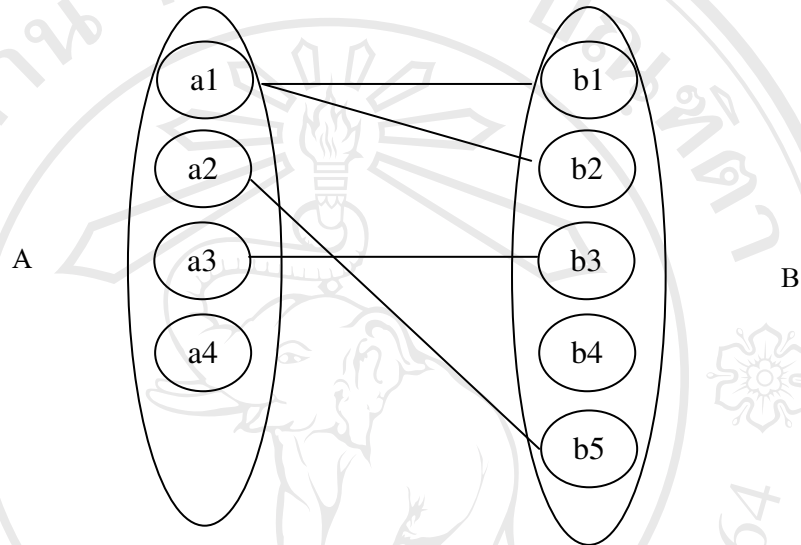
รหัสหนังสือ	รหัสสมาชิก
LB 435	1748
SC 593	1645
GM 185	1540
CO 556	2001
SC 602	1645

รูป 2.4 อธิบายความสัมพันธ์แบบหนึ่งต่อหนึ่งด้วยตารางความสัมพันธ์

จากตัวอย่าง ความสัมพันธ์ดังกล่าว หมายถึง สมาชิกแต่ละคน สามารถที่จะยืมหนังสือได้เพียงคนละ 1 เล่มเท่านั้น และหนังสือเล่มดังกล่าวก็ไม่สามารถให้ใครยืมได้อีก เพราะมีเพียงเล่มเดียว ในที่นี้จะสังเกตได้ว่าทั้ง 2 เอนทิตี จะมีคีย์หรือข้อมูลหลักคือ รหัสสมาชิก

ความสัมพันธ์แบบหนึ่งต่อกลุ่ม

เป็นลักษณะความสัมพันธ์ที่ระเบียนหนึ่งระเบียนในเอนทิตี A มีความสัมพันธ์กับระเบียนหลายระเบียนในเอนทิตี B และในทางกลับกันหลายระเบียนในเอนทิตี B จะมีความสัมพันธ์กับระเบียนเพียงหนึ่งระเบียนในเอนทิตี A



รูป 2.5 แสดงความสัมพันธ์ในลักษณะหนึ่งต่อกลุ่ม

สมาชิก

การยืมหนังสือ

รหัสสมาชิก	ชื่อสมาชิก	แผนก
1540	อรนุช	การขาย
1645	กึ่งแก้ว	เลขานุการ
1748	ภาวนา	การบัญชี
1985	สมนึก	คอมพิวเตอร์

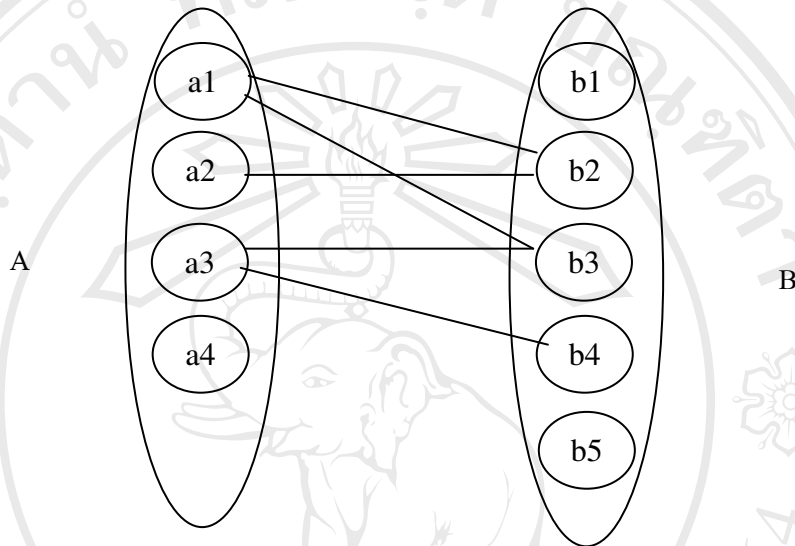
รหัสหนังสือ	รหัสสมาชิก
LB 435	1748
SC 593	1645
GM 185	1540
CO 556	2001
SC 602	1645

รูป 2.6 อธิบายความสัมพันธ์แบบหนึ่งต่อกลุ่มด้วยตารางความสัมพันธ์

จากตัวอย่าง สมาชิกที่มีรหัสสมาชิก 1540 สามารถยืมหนังสือได้ 2 เล่ม คือ รหัสหนังสือ LB 435 และ SC 593 แต่หนังสือทั้ง 2 เล่ม จะถูกยืมด้วยสมาชิกได้เพียงคนเดียว

ความสัมพันธ์แบบกลุ่มต่อกลุ่ม

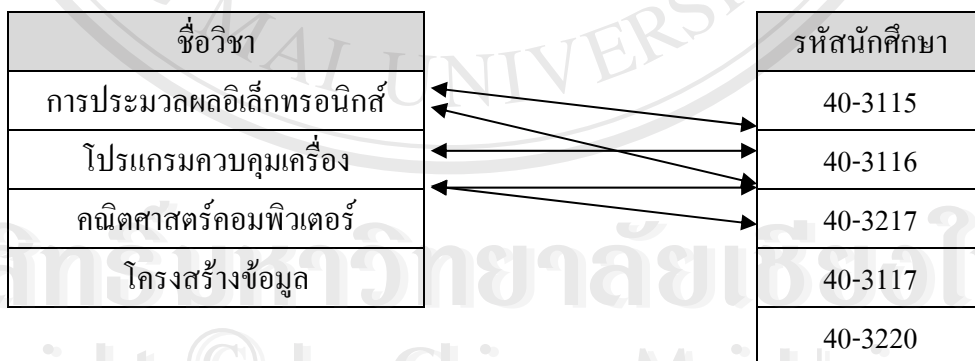
เป็นความสัมพันธ์ที่แต่ละระเบียบในเอนทิตี A จะมีความสัมพันธ์กับระเบียบหลายระเบียบในเอนทิตี B และในทางกลับกันแต่ละระเบียบในเอนทิตี B จะมีความสัมพันธ์กับระเบียบที่อยู่ในเอนทิตี A



รูป 2.7 แสดงความสัมพันธ์ในลักษณะกลุ่มต่อกลุ่ม

วิชา

นักศึกษา



รูป 2.8 อธิบายความสัมพันธ์แบบกลุ่มต่อกลุ่มด้วยตารางความสัมพันธ์

จากตัวอย่าง ในแต่ละรายวิชา จะสามารถมีนักศึกษาเรียนได้มากกว่า 1 คน และในทางกลับกันนักศึกษาแต่ละคนก็สามารถเลือกเรียนวิชาได้มากกว่าหนึ่งรายวิชา

2) รูปแบบของฐานข้อมูล

โครงสร้างของข้อมูลโดยทั่วไปจะมี 3 แบบด้วยกัน คือ

- ฐานข้อมูลแบบลำดับชั้น (Hierarchical Database)
- ฐานข้อมูลแบบเครือข่าย (Network Database)
- ฐานข้อมูลแบบเชิงสัมพันธ์ (Relational Database)

ฐานข้อมูลแบบลำดับชั้น

ลักษณะของโครงสร้างของฐานข้อมูลแบบลำดับชั้นนี้ จะมีลักษณะคล้ายต้นไม้ ที่คำว่าห้วงจึงอาจเรียกโครงสร้างข้อมูลแบบนี้ได้อีกแบบว่าเป็น โครงสร้างแบบต้นไม้ (Tree Structure) โดยจะมีระเบียบที่อยู่ด้านบนซึ่งจะเรียกว่าเป็น ระเบียบพ่อแม่ (Parent Record) ระเบียบในแถวถัดลงมาจะเรียกว่าระเบียบลูก (Child Record) ซึ่งระเบียบพ่อแม่จะสามารถมีระเบียบลูกได้มากกว่าหนึ่งระเบียบ แต่ระเบียบลูกแต่ละระเบียบจะมีพ่อแม่เพียงหนึ่งระเบียบเท่านั้น

ระเบียบพนักงานขาย

รหัสพนักงานขาย	ชื่อพนักงานขาย
----------------	----------------



ระเบียบลูกค้า

รหัสลูกค้า	ชื่อลูกค้า
------------	------------



ระเบียบสินค้า

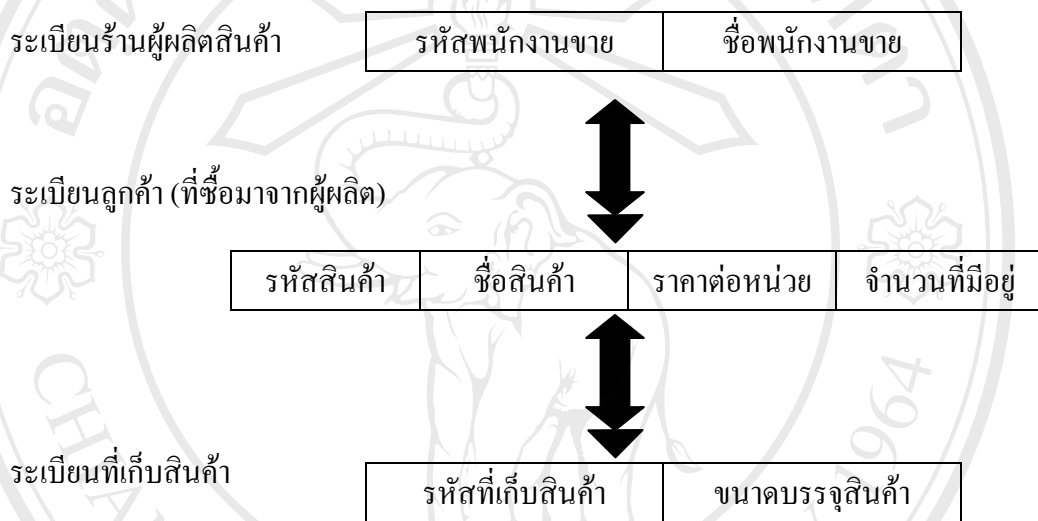
รหัสสินค้า	จำนวนสินค้า
------------	-------------

รูป 2.9 แสดงโครงสร้างฐานข้อมูลแบบลำดับชั้น

จากตัวข้างต้นจะเห็นว่า ลูกค้าแต่ละคนจะไม่สามารถได้รับการจากพนักงานขายมากกว่าหนึ่งคนได้ เนื่องจากลูกค้าแต่ละคนถือว่าเป็นระเบียบลูก และพนักงานขายจะถือว่าเป็นระเบียบพ่อแม่ของลูกค้า สินค้าแต่ละชนิดก็จะถูกซื้อ โดยลูกค้าเพียงคนเดียวเท่านั้น เนื่องจากสินค้าแต่ละชนิด จะเป็นระเบียบลูกของระเบียบลูกค้า เป็นต้น

ฐานข้อมูลแบบเครือข่าย

ข้อมูลภายในฐานข้อมูลแบบนี้สามารถมีความสัมพันธ์กันแบบใดก็ได้ เช่น อาจเป็นแบบหนึ่งต่อหนึ่ง หนึ่งต่อกลุ่ม ตัวอย่างของฐานข้อมูลแบบนี้ เช่น การสั่งซื้อสินค้าจากร้านผู้ผลิตสินค้า และการนำสินค้าไปเก็บในคลังสินค้า ซึ่งจะแสดงความสัมพันธ์ระหว่างทะเบียนร้านผู้ผลิตสินค้าและทะเบียนสินค้า และความสัมพันธ์ระหว่างทะเบียนสินค้าและทะเบียนที่เก็บสินค้าได้ โดยการใช้ลูกศรเชื่อมโยงเช่นกัน ดังรูป 2.10



รูป 2.10 แสดงโครงสร้างฐานข้อมูลแบบเครือข่าย

ฐานข้อมูลแบบเชิงสัมพันธ์

ฐานข้อมูลเชิงสัมพันธ์ จะมีโครงสร้างข้อมูลต่างจากฐานข้อมูลสองแบบแรก คือ ข้อมูลจะถูกเก็บอยู่ในรูปแบบของตาราง (Table) ซึ่งภายในตารางก็จะแบ่งออกเป็นแถว (Row) และ คอลัมน์ (Column) แต่ละตารางจะมีจำนวนแถวได้หลายแถว และจำนวนคอลัมน์ได้หลายคอลัมน์ แถวแต่ละแถวจะสามารถเรียกได้อีกอย่างว่าทะเบียนหรือเรคคอร์ด (Record) คอลัมน์แต่ละคอลัมน์สามารถเรียกได้อีกอย่างว่า เขตข้อมูลหรือฟิลด์ (Field)

2.2 แนวความคิดและหลักการออกแบบเว็บเพจ

วัชรพงศ์ ยะไวทย์ (2544) กำหนดว่า หลักการออกแบบเว็บเพจมี 15 ประการ คือ

- 1) ต้องตัดสินใจให้แน่นอนว่า มีวัตถุประสงค์ในการทำเว็บเพจเพื่ออะไร ซึ่งโดยทั่วไปจะมีสามแบบ คือ เพื่อประชาสัมพันธ์บริษัท หรือเพื่อขายปลีกแบบใส่แคตตาล็อก (Online Catalog) หรือเพื่อแสดงสินค้าเฉพาะที่เด่น ๆ
- 2) ต้องทราบกลุ่มเป้าหมายว่าเป็นใครและต้องการอะไร เช่น ผู้ที่ซื้อดอกไม้ผ่านทางเว็บไซต์ ไม่ได้ต้องการดอกไม้ราคาถูก แต่ซื้อเพราะต้องการบอกความในใจให้แก่คนที่รักและห่วงใย เพราะฉะนั้นการออกแบบเว็บเพจก็ควรจะแสดงภาพของความงามหรือคุณค่า มากกว่าการขายถูก
- 3) ต้องออกแบบให้เข้าถึงข้อมูลที่กลุ่มเป้าหมายต้องการให้เร็วและครบถ้วนที่สุด เช่น การจัดทำหมวดหมู่ของข้อมูล การสามารถส่งผ่านไปยังจุดใดจุดหนึ่งได้ง่าย (Link)
- 4) เว็บไซต์ที่ดีจะต้องโหลด (Load) ข้อมูลได้รวดเร็ว ภายใน 8 – 15 วินาทีแรก ควรจะดึงดูดความสนใจได้ และไม่เกิน 30 วินาทีควรจะโหลดเสร็จ
- 5) ควรมีเมนูเฉพาะหรือที่เป็นชอร์ตคัต (Shortcut) สำหรับหมวดหรือหน้าที่มีผู้เข้าเยี่ยมชมจำนวนมาก
- 6) ควรใช้รูปสัญลักษณ์มากกว่าคำบรรยาย แต่ควรทำให้สมดุลกันทั้งสองอย่าง เพราะรูปมากไปก็ทำให้โหลดช้า
- 7) หากมีภาพจำนวนมากให้จัดทำเป็นภาพเล็ก ๆ และมีคำอธิบายโดยย่อไว้ และทำให้รูปขยายได้ เมื่อผู้เยี่ยมชมเข้าไปคลิกดูภาพขยายหรือรายละเอียดอื่น ๆ ได้
- 8) ควรเปิดทางเลือกให้ผู้เยี่ยมชม ไม่ควรจะบังคับไปในทิศทางใดทางหนึ่งมากเกินไป เช่น ไปเยี่ยมชมหน้าไหนก็เจอแต่แบบฟอร์มลงทะเบียน หากไม่ลงทะเบียนไม่ให้ชม เป็นการปิดกั้นเว็บไซต์จนเกินไป
- 9) ควรออกแบบให้เป็นกันเอง หรือเป็นมิตรกับผู้เยี่ยมชม และควรให้อะไรกับสังคมบ้างเช่น การขายสินค้าลดทอน ก็ควรจะให้ความรู้เกี่ยวกับสิ่งของนั้น ๆ หรือมีเรื่องของประเพณีไทยด้วย
- 10) ในหนึ่งหน้าเว็บเพจไม่ควรจะมีข้อมูลยาวเกิน 3 หน้าจอ หรือ 1 หน้า A4 ทั้งนี้เพื่อประโยชน์ในการโหลดข้อมูลได้เร็ว และช่วยในการจัดหน้าพิมพ์เก็บเข้าแฟ้มด้วย
- 11) การจัดข้อมูลเท็กซ์ในหน้าเว็บเพจควรจะจัดเป็นคอลัมน์ไม่เกิน 500 Pixel (ไม่เกิน 60 – 70 เปอร์เซ็นต์ของความกว้างหน้าจอ) ทั้งนี้เพื่อต่อการอ่าน และไม่ทำให้เบื่อก่อน

12) ไม่ควรใช้วิธีลิงค์ข้อมูลเป็นแบบช่วง ๆ ในหน้าเดียวกัน เพราะนอกจากจะทำให้เสียเวลาในการโหลดข้อมูลแล้ว ก็ยังทำให้ผู้เยี่ยมชมเสียความรู้สึกเมื่อสกรอลล์อ่านลงมาอีกครั้งหนึ่งด้วยความไม่รู้

13) ควรออกแบบเว็บเพจให้สนองตอบต่อทุกเวอร์ชันหรือทุกค่ายของบราวเซอร์

14) ควรออกแบบเว็บเพจให้ง่ายต่อการบำรุงรักษา และควรจะดูแลแก้ไขข้อมูลอยู่เป็นประจำ ซึ่งเป็นปัจจัยที่สำคัญที่จะทำให้มีผู้เข้าเยี่ยมชมประจำ

15) การออกแบบเว็บเพจในรูปแบบใดนั้น จะขึ้นอยู่กับเป้าหมายและแนวนโยบายที่ได้วางไว้แต่ต้น รวมถึงการออกแบบให้สอดคล้องกับพฤติกรรมของกลุ่มเป้าหมายที่เราต้องการ แนวทางการพัฒนาโปรแกรมหรือเว็บเพจนั้นๆ ด้วย

2.3 แนวความคิดเกี่ยวกับประเภทของการให้บริการจดหมายอิเล็กทรอนิกส์

อี-เมลล์แบบ POP เป็นอี-เมลล์อีกชนิดหนึ่ง ที่มีความสะดวกในการทำงานมากเนื่องจากสามารถเช็คอี-เมลล์ได้จากซอฟต์แวร์เช็คอี-เมลล์ใดก็ได้ แต่ผู้ใช้จะต้องทำการติดตั้งใช้งานเอง เช่น การเช็คค่า incoming mail server ค่า outgoing mail server และอื่นๆ ข้อดีของอี-เมลล์แบบนี้ จะสามารถเช็คอี-เมลล์ได้รวดเร็ว สะดวก ไม่มีขยะปนมากับเนื้อหาที่สำคัญไม่ต้องเข้ามาที่เว็บไซต์นี้บ่อยๆ สามารถแทรกไฟล์ได้ และบริการให้ฟรี มาตรฐานของ POP นี้ใช้กำหนดสำหรับการอ่านอี-เมลล์เข้ามาในเครื่องเท่านั้น แต่ถ้าต้องการส่งอี-เมลล์ต้องใช้ SMTP (Simple Mail Transfer Protocol)

อี-เมลล์แบบ IMAP เป็นบริการอี-เมลล์ที่ทำให้การเปิดอี-เมลล์คล้ายกับ Web-Based อี-เมลล์คือต้องใช้โปรแกรมเฉพาะ เรียกอี-เมลล์มาอ่าน เมื่อปิดโปรแกรมแล้วจะไม่ทิ้งอะไรไว้ในเครื่อง ให้ใครมาแอบเปิดดูได้เหมือนแบบ POP ผู้ใช้สามารถเปิดอี-เมลล์จากที่ใดก็ได้ ด้วยโปรแกรมสำหรับอ่านอี-เมลล์จาก IMAP เซิร์ฟเวอร์เท่าที่ทราบ ยังไม่เห็นเซิร์ฟเวอร์ใด ให้บริการ IMAP ฟรีกับคนทั่วไปแต่เห็นเฉพาะในบริษัท สถาบัน หรือองค์กรเท่านั้น และ IMAP เซิร์ฟเวอร์ก็มักที่จะเป็น UNIX หรือ LINUX เพราะไม่ต้องติดตั้งอะไรมาก ก็ใช้ให้บริการได้เลย เพราะเป็นคุณสมบัติพื้นฐานอยู่แล้ว

อี-เมลล์แบบ Web-Based อี-เมลล์เป็นอี-เมลล์แบบที่เรารู้จักกันเช่น hotmail.com yahoo.com และchaiyo.com ก็ยังเป็น Web-Basedอี-เมลล์เพราะผู้ใช้บริการพอใจที่จะให้บริการมากกว่า POP เนื่องจากการอ่าน หรือส่งอี-เมลล์ผู้ใช้บริการจะต้องเข้ามายังเว็บของผู้ให้บริการทุกครั้ง ต่างกับ POP ที่ผู้ใช้บริการ สามารถใช้โปรแกรมดึงอี-เมลล์ไปอ่าน โดยไม่ต้องเข้าเว็บของผู้ให้บริการแต่อย่างใด

อี-เมลล์แบบ TELNET เป็นรูปแบบการรับส่งอี-เมลล์ในสมัยเริ่มต้น การใช้อี-เมลล์ก็จะสามารถแสดงผลเฉพาะตัวอักษร โดยจะต้องใช้โปรแกรม Telnet ที่มากับระบบปฏิบัติการ เป็น

โปรแกรมช่วยติดต่อกับเครื่องเซิร์ฟเวอร์และต้องไปรัน โปรแกรมที่ชื่อว่า Pine ที่เครื่องเซิร์ฟเวอร์ เพื่อทำการเช็คอี-เมลในรูปแบบการแสดงผลแบบตัวอักษรเท่านั้น

(<http://www.thaiail.com/article/mail.htm#type> : 3 กันยายน พ.ศ. 2547)

2.4 แนวความคิดเกี่ยวกับสแปมเมลล์

2.4.1 ความหมายของสแปมเมลล์

ศูนย์ประสานงานการรักษาความปลอดภัยคอมพิวเตอร์ ประเทศไทย ได้ให้ความหมายของสแปมเมลล์ (Spam Mail) หมายถึง การทำให้ผู้หนึ่งหรือ Newsgroup หนึ่งเต็มไปด้วยข้อความที่ไม่เกี่ยวข้องและไม่เหมาะสม

อรรยา สิงห์สงบ ได้ให้คำจำกัดความของสแปมเป็นภาษาอังกฤษว่า “Spam Mail means An Unsolicited Commercial E-mail (UCE) or An Unsolicited Bulk E-mail (UBE) known as Junk e-mail.”

โดยสรุป สแปมเมลล์ หมายถึง จดหมายอิเล็กทรอนิกส์ที่ผู้ส่ง (ซึ่งมักจะไม่มีปรากฏชื่อและที่อยู่ผู้ส่ง) ได้ส่งไปยังผู้รับอย่างต่อเนื่องโดยส่งจำนวนครั้งละมากๆ และมีได้รับความนิยมนจากผู้รับ โดยการส่งสแปมเมลล์นั้นอาจมีวัตถุประสงค์ในเชิงพาณิชย์หรือไม่ก็ได้

2.4.2 ลักษณะทั่วไปของสแปมเมลล์

1) มักเป็นจดหมายโฆษณาเว็บไซต์ หรือการให้บริการอย่างใดอย่างหนึ่ง ซึ่งบริการหลายอย่างนั้น บ่อยครั้งไม่สามารถจะโฆษณาผ่านช่องทางปกติได้โดยสะดวก อย่างเช่น โฆษณาเกี่ยวกับการเสริมความงามทั้งของหญิงและชาย โฆษณาขายซอฟต์แวร์ หนังสือ เป็นต้น

2) ชื่อผู้ส่งมักจะไม่สามารถค้นหาที่มาได้ ด้วยเหตุที่การส่งจดหมายขยะ เสี่ยงต่อข้อกฎหมาย และเสี่ยงต่อเว็บไซต์ของผู้ส่งจะต้องถูกปิดลงจากการละเมิดการใช้งาน

3) มีจุดเชื่อมต่อที่แจ้งให้ผู้รับสามารถถอนชื่อออก เมื่อผู้รับพยายามติดตามจุดเชื่อมต่อไปยังเว็บไซต์ที่ทำการถอดถอนบริการ มักจะไม่สามารถถอดถอนบริการได้ และมักถูกใช้เป็นช่องทางตรวจสอบว่าอี-เมลล์ของนั้นได้ใช้งานอยู่หรือไม่

4) หัวข้อจดหมายขยะมักจะเป็นข้อความที่ดูน่าสนใจ หรือไม่ก็อาจจะหลอกลวงให้ผู้รับเปิดอ่านให้ได้

5) ส่วนเนื้อหาของจดหมายมักมีการใช้สัญลักษณ์พิเศษ ในภาษาอังกฤษอาจมีการใช้สัญลักษณ์คั่นระหว่างคำ เช่น V!agra หรือใช้ช่องว่างคั่นระหว่างคำ เช่น S E X หรืออาจจะใช้

ตัวหนังสือสีขาวที่มองเห็นยากและขนาด Font เล็กมากอยู่เป็นพื้นหลัง เพื่อให้ยากแก่การตรวจจับ โดยใช้โปรแกรมตรวจสอบ

(<http://pages.cpsc.ucalgary.ca/~schuyt/assignments/supporting/overview.pdf> : 21 มกราคม พ.ศ. 2548)

2.4.3 ผลกระทบของสแปมเมลล์

1) ค่าใช้จ่ายจากการส่งสแปมเมลล์ (Cost)

ในการส่งสแปมเมลล์นั้น ผู้ส่งสแปมเมลล์เสียค่าใช้จ่ายน้อยมากเมื่อเปรียบเทียบกับ การส่งเมลล์ขยะทั่วไป เนื่องจากการส่งนั้นมีลักษณะที่ผลัดภาระค่าใช้จ่ายมายังผู้ให้บริการ อินเทอร์เน็ตและผู้ใช้อินเทอร์เน็ต (Cost Shifting to ISP Costs and Consumer Costs) ซึ่งได้แก่ ค่าใช้จ่ายในการดาวน์โหลด (Download cost) ปัญหาเรื่องเนื้อที่ของการใช้ดิสก์ (Disk space) รวมถึงการเสียเวลาของลูกจ้างในการจัดการกับสแปมเมลล์ (Employee time wasted)

ในแง่ของผู้ใช้อินเทอร์เน็ตในฐานะผู้บริโภคนั้น จะเห็นได้ว่าผู้ใช้จะต้องเสียเวลา เพื่อลบสแปมเมลล์ออกจากตู้ไปรษณีย์อิเล็กทรอนิกส์ (Mailbox) ของตน ซึ่งถ้าหากเป็นกรณีที่ใช้ อินเทอร์เน็ตไม่ได้เหมาะจ่ายค่าธรรมเนียมการใช้อินเทอร์เน็ต โดยอาจจ่ายการคำนวณตามเวลาของ การใช้ การลบสแปมเมลล์ทำให้เสียเวลามากและกินเวลาการใช้อินเทอร์เน็ต อันมีผลทำให้ผู้ใช้ อินเทอร์เน็ตต้องเสียค่าธรรมเนียมการใช้อินเทอร์เน็ตเพิ่มขึ้น หรือไม่เช่นนั้น ผู้ใช้อินเทอร์เน็ตก็ต้อง เสียค่าใช้จ่ายเพิ่ม โดยการซื้อเนื้อที่ในการส่งและรับจดหมายเพิ่มขึ้น ถ้าหากตนไม่ต้องการเสียเวลา ในการแยกจดหมายของตนออกจากสแปมเมลล์

สำหรับผู้ให้บริการอินเทอร์เน็ต (ISP) นั้นก็ต้องเสียค่าใช้จ่ายในการจ้างพนักงาน เพิ่ม เพื่อคัดสแปมเมลล์ออกจากระบบ นอกจากนี้ ISP ยังจะต้องขยายความกว้างของช่องสัญญาณ (Bandwidth) เพื่อรักษาระดับความเร็วและการเข้าสู่ระบบ และค่าใช้จ่ายในการเพิ่มระบบ กู้คืนกรองสแปมเมลล์เพื่อรักษาลูกค้าของตนไว้

2) ความเป็นส่วนตัว (Privacy)

การรวบรวมชื่อและที่อยู่ของผู้ใช้ไปรษณีย์อิเล็กทรอนิกส์ของผู้ส่งสแปมเมลล์ โดย ผู้เป็นเจ้าของชื่อและที่อยู่ไม่นั้นไม่ทราบและไม่ได้ให้ความยินยอมนั้นถือเป็นการละเมิดความเป็นส่วนตัว

3) การหลอกลวงโดยส่งจดหมายซึ่งใช้ที่อยู่หรือหัวเรื่องของไปรษณีย์อิเล็กทรอนิกส์ปลอม (Spoofing)

การส่งจดหมายโดยใช้ที่อยู่หรือหัวเรื่องของไปรษณีย์อิเล็กทรอนิกส์ปลอมซึ่งบางครั้งก็ใช้ชื่อและที่อยู่หน่วยงานที่มีชื่อเสียงเพื่อที่จะหลอกให้ผู้รับสนใจที่จะเปิดจดหมายของตนหรือตอบกลับ

4) เนื้อหาของจดหมายที่มีลักษณะหลอกลวงและขัดต่อกฎหมายและศีลธรรมอันดี (Content Fraudulent & Deceptive Advertising)

นอกจากนี้การโฆษณาสินค้าของสแปมเมอร์ส่วนใหญ่มีลักษณะฉ้อฉล เช่น โฆษณาว่ามีการแจกสินค้าฟรีหากส่งจดหมายนี้ต่อไปยังคนที่รู้จักอย่างน้อย 20 คน หรือมีการโฆษณาที่มีเนื้อหาขัดต่อกฎหมายหรือศีลธรรมอันดี เช่น การโฆษณาเว็บไซต์ภาพลามก การพนัน เป็นต้น

5) การบุกรุก (Trespass)

มีการโต้แย้งกันว่าการส่งสแปมเมอร์ถือเป็นการบุกรุกสิทธิความเป็นส่วนตัวของผู้รับสแปมเมอร์ ถึงแม้ว่าสแปมเมอร์มักจะให้สิทธิแก่ผู้รับสแปมเมอร์ เลือกที่จะคลิกยกเลิกมิให้มีการส่งสแปมอีกต่อไป (A Remove from List) การให้สิทธิดังกล่าวไม่ได้ผลเนื่องจากการกระทำดังกล่าวกลับเป็นการยืนยันกับผู้ส่งสแปมเมอร์ว่า ชื่อและที่อยู่ดังกล่าวนั้นมีอยู่จริงและผู้ส่งสแปมเหล่านี้จะนำชื่อและที่อยู่ดังกล่าวไปสร้างลิสต์ใหม่เพื่อจำหน่ายให้กับผู้ส่งสแปมเมอร์รายต่อ ๆ ไป

2.4.4 เทคนิคการกำจัดสแปมเมอร์

ในปัจจุบันได้มีการพัฒนาเทคนิคในการกำจัดสแปมเมอร์ขึ้นมาหลากหลายวิธี เพื่อให้สามารถตรวจจับสแปมเมอร์ได้อย่างมีประสิทธิภาพมากขึ้น เนื่องจากสแปมเมอร์เองก็ได้คิดค้นวิธีการสแปมใหม่ ๆ เพื่อหลีกเลี่ยงจากการตรวจจับสแปมเช่นกัน ทางผู้วิจัยได้ทำการศึกษาค้นคว้ารวบรวม และสรุปเทคนิควิธีในการกำจัดสแปมเมอร์ต่าง ๆ ออกเป็น 7 วิธีคือ

1) ร่วมด้วยช่วยกันแบล็คลิสต์

มีการรวมตัวขององค์กรที่ช่วยกันรวบรวมรายชื่ออีเมลเซิร์ฟเวอร์ ที่มีพฤติกรรมส่งสแปมเมอร์ หรือเปิดช่องโหว่ในระบบรักษาความปลอดภัย จนเป็นช่องทางให้แฮกเกอร์เข้ามาใช้ส่งสแปมเมอร์ได้ แล้วตั้งเป็นแบล็คลิสต์สากล เพื่อแจกจ่ายข้อมูลให้เมลเซิร์ฟเวอร์ที่ร้องขอรายชื่อทั่วโลก แต่วิธีนี้ยังไม่ค่อยได้ผลนัก เพราะรายชื่อที่มีการแบล็คลิสต์ไว้ มีเพียงน้อยนิดเมื่อเทียบกับจำนวนของเมลเซิร์ฟเวอร์ ที่มีการส่งสแปม ที่เกิดขึ้นใหม่ทุกวัน อีกทั้งองค์กรที่ทำแบล็คลิสต์บางราย ใช้วิธีที่ไม่เหมาะสม จนทำให้เมลเซิร์ฟเวอร์ดี ๆ ไปติดอยู่ในแบล็คลิสต์ด้วยเช่นกัน

2) ใช้สแปม-แพทเทิร์น (Spam-Pattern) แบบเดียวกับไวรัส

วิธีสแปม-แพทเทิร์นนี้ เป็นการนำสแปมเมอร์ที่เคยพบ มาคำนวณหารูปแบบเฉพาะตัว หรือที่เรียกว่า “ซิกเนเจอร์” แล้วรวบรวมเป็นฐานข้อมูลเพื่อใช้ในการตรวจสอบกับอี-เมล

ฉบับอื่น ๆ ที่ผ่านเข้ามาในระบบ หากมีการตรวจพบว่าอี-เมลล์ฉบับใด ตรงกับซิกเนเจอร์ที่รวบรวมไว้ ก็แสดงว่าอี-เมลล์ฉบับนี้เป็นสแปมเมลล์

สแปมเมลล์นั้น สามารถสร้างได้ง่ายและรวดเร็วกว่าไวรัสมาก แม้แต่บุคคลทั่วไป ที่ไม่มีความรู้ทางด้านเทคนิคมากนัก ก็สามารถส่งสแปมเมลล์จำนวนมากได้ รูปแบบของสแปมเมลล์จึงมีความหลากหลาย และยากต่อการผลิตซิกเนเจอร์ให้ทันต่อการสแปมที่เกิดขึ้น และทำให้ดาตาเบสของซิกเนเจอร์มีขนาดใหญ่ ถ้าหากอี-เมลล์เข้าทุกฉบับที่ได้รับ จะต้องผ่านการเปรียบเทียบกับซิกเนเจอร์จำนวนกว่าล้านครั้ง ประสิทธิภาพของระบบอี-เมลล์จะลดลงมากเพียงใด

3) การคำนวณเบย์เซียน

วิธีนี้เป็นที่ได้รับความนิยมรับทั่วโลกว่าดีที่สุดในปัจจุบัน หลักการที่ใช้ก็คือ การวิเคราะห์อี-เมลล์ตัวอย่างจำนวนมหาศาลที่แบ่งเป็น 2 กลุ่ม คือ กลุ่มอี-เมลล์ปกติ และกลุ่มอี-เมลล์ที่เป็นสแปม เพื่อวิเคราะห์คำศัพท์ที่ปรากฏอยู่ในอี-เมลล์ทั้งสองกลุ่ม หากคำศัพท์คำใด มีปรากฏอยู่ในสแปมเมลล์เป็นจำนวนมาก อี-เมลล์ที่มีคำ ๆ นั้นอยู่ ก็มีความน่าจะเป็นที่จะเป็นสแปมเมลล์สูงนั่นเอง

ระบบเบย์เซียนมีความแม่นยำสูงมากถึง 99.9 เปอร์เซ็นต์ แต่ข้อเสียของวิธีนี้ก็มีส่วน เช่นกัน คือ การจะได้มาซึ่งค่าความน่าจะเป็นที่เหมาะสม จะต้องรวบรวมอี-เมลล์ที่เป็นสแปมและไม่ใช่สแปมจำนวนมากเพื่อเทรนระบบ และให้การตัดสินใจของระบบเหมาะสมกับสภาพแวดล้อมของผู้ใช้งาน เช่น ในประเทศไทย การใช้ระบบที่ถูกเทรนมาจากต่างประเทศ ย่อมไม่สามารถแยกแยะอี-เมลล์ที่เป็นภาษาไทยได้ หรือหน่วยงานที่ดำเนินงานทางธุรกิจ กลุ่มของคำในอี-เมลล์ที่ไม่เป็นสแปม ก็จะต้องเกี่ยวกับการทำธุรกิจ การติดต่อค้าขาย ในขณะที่ คำเหล่านั้น มักจะเป็นสแปมในกลุ่มผู้ใช้งานอี-เมลล์ที่เป็นนักเรียน นักศึกษา นักวิจัย และอาจารย์ในมหาวิทยาลัย เป็นต้น

4) การกรองฐานกฎ (Rule-Base Filtering)

วิธีนี้เป็นการกำหนดคะแนนความเป็นสแปมให้กับ คำ วลี และ คุณสมบัติต่าง ๆ ของอี-เมลล์ แล้วตั้งเป็นกฎขึ้นมาตรวจสอบกับอี-เมลล์ทุกฉบับ เช่น หากมีคำว่า “ขายตรง” คะแนนก็จะ เป็น +3 ถ้ามีการใช้ภาษาเอชทีเอ็มแอล (HTML) คะแนนก็จะ เป็น +1 ถ้ามีการใช้ตัวอักษรสีแดง หรือตัวอักษรขนาดใหญ่ คะแนนก็จะ เป็น +2 เป็นต้น

เมื่อมีเมลล์ผ่านเข้ามาในระบบก็จะถูกตรวจสอบว่าตรงกับกฎข้อใดบ้าง ระบบจะรวมคะแนนที่ได้จากกฎที่ตรวจพบทั้งหมด แล้วเปรียบเทียบกับคะแนนที่ใช้เป็นเกณฑ์ในการตัดสินใจ หากมีค่ามากเกินเกณฑ์ที่กำหนด ก็จะถูกระบุว่าเป็นสแปม

วิธีการนี้เป็นวิธีการที่นิยมใช้กันมากที่สุดในปัจจุบัน เพราะเข้าใจง่าย ติดตั้งง่าย และมีความแม่นยำสูงพอสมควร (ประมาณ 90 เปอร์เซ็นต์) แต่วิธีนี้มีข้อเสียคือ อี-เมลล์ที่ไม่ใช่สแปม มักจะถูกตัดสินเป็นสแปมบ่อย ๆ เพราะคะแนนของกฎที่กำหนดไว้ ไม่ได้เหมาะสมกับอี-เมลล์ทุก

ประเภท ข้อเสียอีกประการหนึ่งคือ กฎที่กำหนดไว้มีค่าคะแนนคงที่ตายตัว และสแปมเมอร์ก็สามารถหาซอฟต์แวร์เหล่านั้นมาใช้งานได้เช่นกัน เมื่อสแปมเมอร์ต้องการจะส่งสแปม ก็ทำการทดสอบสแปมเมลล์ของตนเสียก่อน และปรับแก้เนื้อหาจนสามารถส่งผ่านระบบการกรองฐานกฎได้

5) ชาเลนจ์-เรสปอนส์ ฟิลเตอร์ริง (Challenge-Response Filtering)

วิธีนี้เป็น การเพิ่มขั้นตอนในการยืนยันตัวเองให้กับการส่งจดหมาย โดยเมื่อมีการส่งจดหมายไปยังผู้รับที่มีการป้องกันด้วยระบบชาเลนจ์-เรสปอนส์ ฟิลเตอร์ริง ระบบจะตรวจสอบว่าผู้ส่งเป็นผู้ที่อยู่ในไวท์ลิสต์ใช่หรือไม่ ถ้าใช่ก็จะส่งจดหมายไปถึงผู้รับตามปกติ แต่ถ้าไม่ใช่ ระบบจะทำการระงับการส่งอี-เมลล์ฉบับนั้นไว้ชั่วคราว แล้วส่งอี-เมลล์ตอบกลับไปยังผู้ส่ง ให้เข้ามายืนยันตัวเอง โดยการเข้ามากรอกแบบฟอร์มที่ระบบเตรียมไว้ให้ ระบบจึงจะส่งอี-เมลล์ฉบับนั้นไปยังผู้รับ

วิธีนี้ก่อให้เกิดความลำบากต่อสแปมเมอร์ผู้ส่งอี-เมลล์ไปให้ผู้รับนับพันนับหมื่นฉบับ เนื่องจากไม่สามารถเสียเวลากับการกรอกแบบฟอร์มเพื่อส่งอี-เมลล์ตอบกลับให้ผู้รับรายใดรายหนึ่งได้ และเมื่อไม่มีการกรอกแบบฟอร์มภายในระยะเวลาที่กำหนด ระบบก็จะดรอปอี-เมลล์ฉบับนั้นทิ้งไป วิธีนี้จึงสามารถป้องกันสแปมเมลล์ได้ในอัตราสูงเกือบ 100 เปอร์เซ็นต์ แต่ก็เป็นการเพิ่มขั้นตอนในการรับส่งอี-เมลล์ที่ไม่เป็นสากล และผู้ใช้งานอินเทอร์เน็ตทั่วไปจำนวนมากไม่รู้จักระบบนี้ จึงทำให้ข่าวสารสำคัญ ๆ ที่ส่งถึงผู้รับ มีความเสี่ยงที่จะสูญหายหรือได้รับอี-เมลล์ฉบับนั้นช้าเกินไปก็ได้ ในบางครั้งวิธีชาเลนจ์-เรสปอนส์ ฟิลเตอร์ริงนี้จึงถูกนำไปใช้ร่วมกับวิธีอื่น ๆ เช่น วิธีเบย์เซียน โดยระบบจะทำการชาเลนจ์-เรสปอนส์ เฉพาะอี-เมลล์ที่ถูกตัดสินว่าเป็นสแปม จากวิธีเบย์เซียนเท่านั้น

6) เอฟเอฟบี (FFB: Filters that Fight Back)

สแปมเมลล์จำนวนมากมีเป้าหมายเพื่อการโฆษณา จะมีการสอดแทรกลิงค์ของเว็บไซต์ที่ขายสินค้าและบริการนั้น หากผู้ได้รับสแปมเมลล์ทุกคนเข้าไปเยี่ยมชมเว็บไซต์เหล่านั้นพร้อมกันทุกคน จะทำให้เซิร์ฟเวอร์ของเว็บไซต์เหล่านั้นต้องรับภาระหนักจนอาจต้องหยุดให้บริการ

แนวความคิดนี้มีผู้นำไปคิดและพัฒนาต่อจนกลายเป็นระบบเอฟเอฟบี หากระบบกรองสแปมเมลล์มีความสามารถในการโจมตีเว็บเหล่านั้นโดยอัตโนมัติ โดยผู้รับจดหมายไม่ได้รับผลกระทบใด ๆ ก็จะเป็นการทำให้เซิร์ฟเวอร์ของสแปมเมอร์รับภาระหนักเกินไป จนต้องหยุดให้บริการ ดังนั้นผู้ที่ต้องส่งสแปม จะต้องเพิ่มประสิทธิภาพของเครื่องที่ใช้งานและระบบเครือข่าย เพื่อรองรับการโจมตีของระบบเอฟเอฟบี ซึ่งนั่นหมายถึงค่าใช้จ่ายที่เพิ่มขึ้น

ข้อเสียของวิธีนี้อยู่ตรงที่ การตอบโต้กลับเช่นนี้ เป็นการโจมตีระบบเครือข่ายอย่างหนึ่ง ซึ่งผู้ใช้งานบางคนเห็นว่าเป็นสิ่งที่ไม่ควรทำ แม้ว่าเป้าหมายจะเป็นสแปมเมอร์ก็ตาม

7) แอดเดรสการ์ด (AddressGuard™)

วิธีนี้เป็นวิธีการป้องกันสแปมเมลที่คิดค้นโดย Yahoo.com และใช้ได้กับผู้สมัครสมาชิกอี-เมลล์ของ Yahoo เท่านั้น โดยมีหลักการป้องกันสแปมเมลล์ คือ สร้างชื่ออี-เมลล์แอดเดรสปลอมขึ้นมา โดยชื่ออี-เมลล์แอดเดรสปลอมนี้จะไม่บ่งบอกถึงชื่อ หรืออี-เมลล์แอดเดรสจริงของผู้ใช้ ชื่ออี-เมลล์แอดเดรสปลอมนี้ สามารถสร้างได้ถึง 500 ชื่อ ถ้าผู้ใช้จำเป็นต้องให้ชื่ออี-เมลล์แอดเดรสบนอินเทอร์เน็ต ผู้ใช้สามารถให้ชื่ออี-เมลล์แอดเดรสปลอมเหล่านี้ได้ เมื่อมีอี-เมลล์ส่งมาถึงผู้ใช้ที่ชื่ออี-เมลล์แอดเดรสปลอม ระบบจะทำการส่งต่ออี-เมลล์จากชื่ออี-เมลล์แอดเดรสปลอมไปยังอี-เมลล์แอดเดรสจริงของผู้ใช้ แต่ถ้าเริ่มมีการส่งสแปมเมลล์มายังชื่ออี-เมลล์แอดเดรสปลอมนี้เมื่อใด ระบบจะจัดการยกเลิกอี-เมลล์ปลอมนั้นอัตโนมัติ โดยอี-เมลล์แอดเดรสจริงของผู้ใช้จะยังคงใช้งานได้ตามปกติ

โดยสรุป วิธีนี้ถือเป็นการป้องกันสแปมเมลล์โดยตัดช่องทางการส่งสแปมเมลล์ซ้ำ ๆ จำนวนมากได้ แต่ไม่ได้ช่วยคัดกรองสแปมเมลล์ในตู้จดหมายของผู้ใช้ จัดเป็นการป้องกันที่ต้นเหตุมากกว่าปลายเหตุ ซึ่งอาจจะต้องใช้วิธีอื่น ๆ ดังที่กล่าวมาแล้วร่วมกัน เพื่อให้การป้องกันสแปมเมลล์มีประสิทธิภาพมากยิ่งขึ้น

จากการศึกษาวิธีป้องกัน คัดกรองสแปมเมลล์ทั้ง 7 วิธี ผู้วิจัยได้เลือกใช้วิธีซาลอนจ์-เรสปอนส์ ฟิลเตอร์เป็นวิธีหลักในการป้องกันสแปมเมลล์ เนื่องจากสามารถคัดกรองสแปมเมลล์ได้เกือบ 100 เปอร์เซ็นต์ มีความแม่นยำสูง ผิดพลาดน้อยมากเมื่อเทียบกับวิธีอื่น และเพิ่มความสามารถของระบบโดยอนุญาตให้จดหมายอิเล็กทรอนิกส์ที่ใช้โดเมนอี-เมลล์แอดเดรสของมหาวิทยาลัยเชียงใหม่เข้ามาสู่ตู้จดหมายของผู้ใช้งานระบบได้ โดยไม่จำเป็นต้องยืนยันตัวตนกลับมายังระบบ และมีหน้าจอบอกให้ผู้ใช้งานสามารถเข้าไปตรวจสอบจดหมายที่ระบบทำการเพนดิง (Pending) ไว้ได้ สามารถอ่านเนื้อความในจดหมาย วัน เวลาที่ส่งจดหมายมา และมีลิงค์ให้ผู้ใช้งานเลือกว่าจะให้ชื่ออี-เมลล์แอดเดรสจากผู้ส่งนี้อยู่ในแบล็คลิสต์ หรือไวท์ลิสต์ได้เอง ทำให้ผู้ใช้ไม่พลาดอี-เมลล์สำคัญ ๆ และเป็นการตรวจสอบซ้ำ (Double Check) ทั้งจากระบบ และจากผู้ใช้งานอีกด้วย