# บทที่ 3
# ความรู้พื้นฐานที่เกี่ยวข้องกับงานวิจัย

ในบทนี้จะกล่าวถึงองค์ความรู้ที่เป็นพื้นฐานในงานวิจัยนี้ ซึ่งได้แก่ ทฤษฎีที่เกี่ยวข้องกับอีเมล์ องค์ประกอบของอีเมล์ ความรู้พื้นฐานเกี่ยวกับทฤษฎีเบย์เซียน (Bayesian Theorem) ความรู้พื้นฐานเกี่ยวกับแขนงการตัดสินใจ (Decision Tree) และความรู้พื้นฐานเกี่ยวกับกระบวนการเจเนติกอัลกอริทึม (Genetic Algorithm)

## 3.1 ทฤษฎีที่เกี่ยวข้องกับอีเมล์

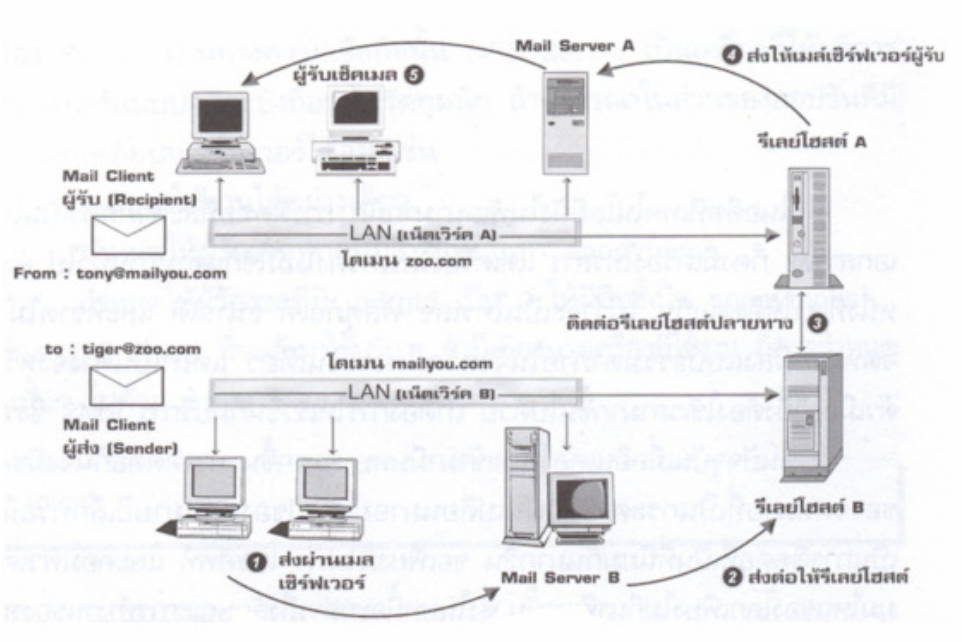
### 3.1.1 ความเป็นมาของอีเมล์

บทความ[14] ได้กล่าวว่า อีเมล์หรือจดหมายอิเล็กทรอนิกส์ ได้มีการใช้งานมานานแล้ว ตั้งแต่ยุคของเครื่องเมนเฟรมหรือมินิคอมพิวเตอร์ ซึ่งไอบีเอ็มได้พัฒนาระบบที่เรียกว่า PROFS (Professional Office System) ออกมาใช้งาน นอกจากนี้ยังมีบนระบบ UNIX อีกด้วย ต่อมาหลายค่ายก็ได้พัฒนาระบบอีเมล์ของตนขึ้นมา โดยส่วนใหญ่จะเป็นองค์ประกอบในแอพพลิเคชั่นที่ทำงานบนระบบเครือข่าย เช่น ไมโครซอฟต์เมล์(Microsoft Mail) ของไมโครซอฟต์, ซีซีเมล์ (CC Mail) ของโลตัส ซึ่งต่างก็ได้ใช้เทคโนโลยีของตนเองและเป็นระบบปิด ดังนั้นการส่งเมล์ไปยังผู้ใช้ที่มีระบบเมล์คนละค่ายกันจึงเป็นเรื่องที่ยุ่งยาก ในยุคต่อมาได้มีระบบเครือข่าย แลน (LAN) และแวน (WAN) ซึ่งต่างมีมาตรฐานและเป็นระบบเปิดมากขึ้น จึงมีการปรับเปลี่ยนการทำงานของระบบเมล์มาเป็นแบบไคลแอนท์-เซิร์ฟเวอร์ (Client-Server) ที่เป็นพื้นฐานที่ใช้กันในระบบยูนิกซ์ (Unix) และมีการพัฒนาอีเมล์เซิร์ฟเวอร์ (Email Server) ทั้งในรูปแบบที่เป็นการใช้งานผ่านระบบ LAN หรือ ใช้งานผ่าน modem ที่เชื่อมต่ออยู่กับระบบ WAN ทำให้ผู้ใช้ไม่สามารถมองเห็นและเข้าถึงไฟล์ในฮาร์ดดิสก์บนเครื่องเซิร์ฟเวอร์ได้ดังนั้นความปลอดภัยของระบบจึงมีมากขึ้น จนในปัจจุบัน ได้พัฒนาขึ้นมาเป็นระบบเวิร์คโฟล์ว (Workflow) ที่ใช้อีเมล์เป็นพื้นฐาน

### 3.1.2 ลักษณะการทำงานของระบบรับ-ส่งอีเมล์

การรับ-ส่งอีเมล์จะมีลักษณะดังรูปที่ 3.1 โดยเริ่มจากผู้ส่ง(Sender) ทำการสร้าง หรือ เขียนอีเมล์ขึ้นมาตามวัตถุประสงค์ที่ต้องการ จากนั้นทำการกดปุ่มส่งโดยผ่านโปรโตคอล(Protocol) ส่งไปยังเครื่องเมล์เซิร์ฟเวอร์ต้นทาง ดังขั้นตอนที่ 1จากนั้นเมล์เซิร์ฟเวอร์จะส่งไปยังเครื่องที่เป็นรีเลย์โฮสต์ (Relay Host) ต้นทาง เนื่องจากรีเลย์โฮสต์เป็นเครื่องที่สามารถติดต่อกับระบบเครือข่ายภายนอกได้ ดังขั้นตอนที่ 2 (โดยทั่วไปเมล์เซิร์ฟเวอร์อาจทำหน้าที่เป็นรีเลย์โฮสต์ในเครื่องเดียวกันก็ได้ ถ้าเป็นเช่นนี้ก็จะไม่มีขั้นตอนที่ 2 เกิดขึ้น) จากรีเลย์โฮสต์ต้นทางเมื่อได้รับเมล์มาแล้ว จะติดต่อ

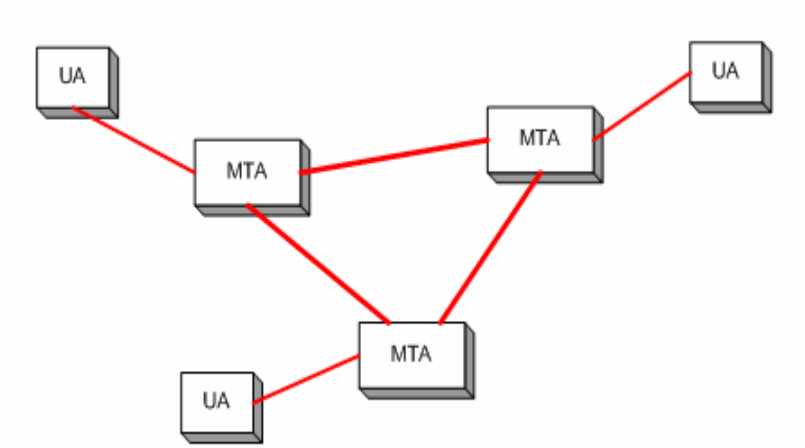
กับรีเลย์โฮสต์ปลายทางเพื่อส่งเมลฉบับนี้ไปดังขั้นตอนที่ 3 และในกรณีเดียวกัน ถ้าเครื่องรีเลย์โฮสต์ปลายทางกับเครื่องเมล์เซิร์ฟเวอร์ปลายทางเป็นเครื่องดียวกันจะไม่มีขั้นตอนที่ 4 เมื่อเมล์ไปถึงเมล์เซิร์ฟเวอร์ปลายทางเรียบร้อยแล้วนั่นคือจบกระบวนการส่งเมล์ เมื่อผู้รับเช็คเมล์ ไม่ว่าจะใช้วิธีไหนก็ตามก็จะต้องติดต่อกับเครื่องเมล์เซิร์ฟเวอร์ของตนเอง เพื่อนำเมล์ฉบับนั้นมาอ่าน ในที่นี้เครื่องเมล์เซิร์ฟเวอร์หรือเครื่องรีเลย์โฮสต์ก็จะทำหน้าที่เหมือนกับที่ทำการไปรษณีย์นั่นเอง




**รูปที่ 3.1** แสดงลักษณะการทำงานของระบบอีเมล์

### 3.1.3 สถาปัตยกรรมของระบบเมล์

ทีซีพี/ไอพี (TCP/IP) มีโปรโตคอลที่สนับสนุนการรับ- ส่งอีเมล์หลายโปรโตคอล แต่โปรโตคอลที่นิยมใช้ในอินเตอร์เน็ตคือ SMTP (Simple Mail Transport Protocol) หน้าที่ของ SMTP คือการกำหนดกรรมวิธีและแบบแผนการนำส่งข้อความระหว่างผู้รับและผู้ส่ง โดย SMTP อาศัยทีซีพีเพื่อลำเลียงจดหมายผ่านพอร์ต 25 ระบบเมล์ที่ใช้ในทีซีพี/ไอพี มีองค์ประกอบสองส่วนคือตัวแทนผู้ใช้ (User Agent :UA) หรืออาจจะเรียกว่าตัวแทนผู้ใช้เมล์ (Mail User Agent : MUA) และตัวแทนขนส่งเมล์ (MTA :Mail Transfer Agent) ทั้งตัวแทนผู้ใช้และตัวแทนขนส่งเมล์เป็นชื่อที่นำมาจากระบบ X.400 ซึ่งเป็นมาตรฐานที่นานาชาติกำหนดไว้เพื่อการนำส่งอีเมล์




**รูปที่ 3.2** แสดงสถาปัตยกรรมในทีซีพี/ไอพี

จากรูปที่ 3.2 ตัวแทนผู้ใช้ทำหน้าที่ในการติดต่อกับผู้ใช้เพื่อรับและส่งอีเมล์ ซึ่งรูปแบบของการติดต่อมี 3 แบบ ดังนี้

1. การติดต่อโดยตรงหรือประมวลผลบนเครื่องที่เก็บเมล์บ๊อกซ์ (Mailbox) อยู่เลย ซึ่งโปรแกรมที่ใช้ในการับส่งเมล์ที่นิยมกันบน Linux/Unix ก็เช่น /bin/mail, mailx, pine, elm เป็นต้น โดยการใช้งานจริงอาจจะด้วยการเทลเนต (Telnet) จากเครื่องคอมพิวเตอร์ส่วนบุคคลเข้าไปยังเครื่องที่เป็นเมล์เซิร์ฟเวอร์แล้วใช้งานโปรแกรมดังกล่าวบนเมล์เซิร์ฟเวอร์

2. การทำงานแบบไคล์แอนท์-เซิร์ฟเวอร์โดยเครื่องที่เป็นเมล์ไคล์แอนท์จะติดต่อกับเครื่องเมล์เซิร์ฟเวอร์โดยผ่านโปรโตคอลสำหรับการจัดการโดยเฉพาะ เช่น POP3 (Post Office Protocol version 3) หรือ IMAP 4 (Internet Mail Access Protocol version 4) เพื่อให้ดึงเมล์จากเมล์บ๊อกซ์บนเซิร์ฟเวอร์ไปอ่านได้โดยสะดวกซึ่งโปรแกรมที่นิยมใช้งานเป็นเมล์ไคล์แอนท์ ได้แก่ Microsoft Outlook, Outlook Express, Endora, Netcape Mail เป็นต้น

3. การทำงานแบบเว็บเมล์ (Web Mail) เป็นการติดต่อระหว่างเว็บเซิร์ฟเวอร์ (Web Server) ที่มีโปรแกรมเว็บเมล์ติดตั้งอยู่กับเมล์เซิร์ฟเวอร์ผ่านโปรโตคอล ที่นิยมใช้กันส่วนใหญ่จะเป็น IMAP ซึ่งเว็บเซิร์ฟเวอร์กับเมล์เซิร์ฟเวอร์อาจจะเป็นเซิร์ฟเวอร์ตัวเดียวกันหรือคนละตัวกันก็ได้ โดยโปรแกรมที่เป็นเว็บเมล์ก็เช่นโปรแกรมที่ติดตั้งอยู่บนเว็บเซิร์ฟเวอร์ของ yahoo.com , hotmail.com เป็นต้น หรือถ้าเป็นโปรแกรมแบบฟรีก็เช่น Horde mail (www.horde.org), OpenWebmail (www.openwebmail.org) , SquirrelMail (www.squirrelmail.org ) เป็นต้น โปรแกรมที่เป็นตัวแทนผู้ใช้ในแบบนี้ก็จะหมายถึงเบราเซอร์ (Browser) ที่รันอยู่บนเครื่องคอมพิวเตอร์ส่วนบุคค ที่ใช้ติดติดต่อไปยังเมล์เซิร์ฟเวอร์ผ่านเว็บเซิร์ฟเวอร์เพื่อดำเนินการในส่วนของการรับและส่งเมล์ ตัวแทนผู้ใช้แบบนี้จะต่างกับแบบที่ 2 คือไม่ต้องมีการใช้โปรโตคอล POP และ IMAP เพราะจะมีตัวกลางที่เป็นเว็บเซิร์ฟเวอร์เป็นตัวใช้งานโปรโตคอลดังกล่าวแทน อาจจะ

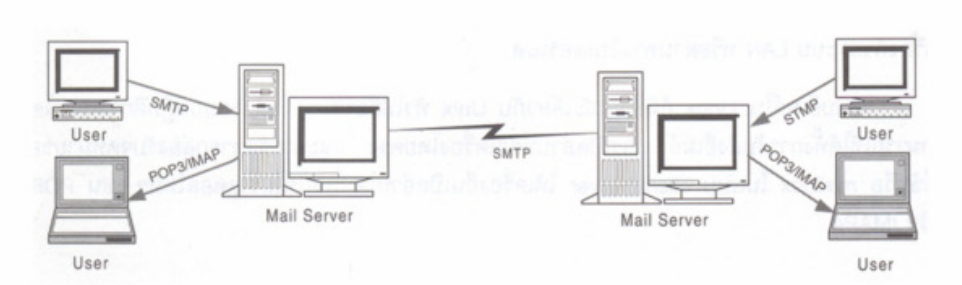
กล่าวได้ว่าตัวแทนผู้ใช้เป็นโปรแกรมอำนวยความสะดวกให้ผู้ใช้เขียน แก้ไข และส่งเมล์ รวมทั้งการเปิดอ่านเมล์ที่ได้รับ และจัดเก็บเมล์เพื่อนำมาใช้ภายหลัง

ตัวแทนขนส่งเมล์ คือส่วนที่ทำหน้าที่ในการรับและส่งอีเมล์ โดยจะรับจากตัวแทนผู้ใช้แล้วตรวจสอบว่าผู้รับปลายทางอยู่ในเครื่องเดียวกันหรือไม่ หากอยู่ในเครื่องเดียวกันก็จะส่งเมล์นั้นไว้ในเมล์บ๊อกซ์หรือโฟลเดอร์ที่เก็บเมล์ของผู้รับนั้น แต่หากอยู่คนละเครื่องจะส่งให้กับอีกกระบวนการหนึ่งเพื่อส่งต่อไปยังเครื่องนั้น ๆ ได้ต่อไป (กระบวนการที่ทำหน้าที่รับส่งเมล์ข้ามเครื่องนั้นอาจเป็น smtpd ที่ทำหน้าที่คอยแปลงเมล์ให้อยู่ในรูปของโปรโตคอล SMTP เพื่อให้สามารถส่งผ่านเครือข่ายทีซีพี/ไอพีได้) ในขณะเดียวกันก็ทำหน้าที่รับเมล์ที่ส่งเข้ามายังผู้รับในเครื่องนั้น แล้วทำการจัดส่งให้ผู้รับแต่ละคนอย่างถูกต้องด้วย ในส่วนนี้โปรแกรมที่นิยมกันก็เช่น Sendmail, Microsoft Mail, Microsoft Exchange

การจัดแบ่งออกแป็นตัวแทนผู้ใช้และตัวแทนขนส่งเมล์มีข้อดีคือ แยกงานของทั้งสองส่วนให้เป็นอิสระจากกัน หน้าที่ของตัวแทนผู้ใช้เน้นการทำงานกับผู้ใช้เพื่อให้ผู้ใช้อ่านเขียนเมล์ได้อย่างสะดวกโดยไม่ต้องยุ่งเกี่ยวกับการทำงานระดับล่างของโปรโตคอล ตัวแทนขนส่งเมล์ทำงานตาม SMTP เช่นการตรวจสอบความถูกต้องของที่อยู่ผู้รับผู้ส่ง รวมทั้งการหาเส้นทางและนำส่งเมล์ไปยังปลายทาง

### 3.1.4 โปรโตคอล

การที่เครื่องคอมพิวเตอร์ 2 เครื่องจะรับส่งเมล์กันได้ หรือผู้ใช้จะโหลดเมล์ไปอ่านที่เครื่องของตนเองนั้น จำเป็นต้องมีโปรโตคอลที่ใช้คุยกันระหว่างเครื่องทั้งสองคือ SMTP POP3 หรือ IMAP ดังรูปที่ 3.3




**รูปที่ 3.3** แสดงโปรโตคอลที่ใช้ในการรับส่งเมล์

#### 3.1.4.1 SMTP

SMTP หรือ Simple Mail Transfer Protocol เป็นโปรโตคอลที่ติดต่อกันระหว่างเครื่องที่เป็นโฮสต์กับโฮสต์ โดยโฮสต์ในที่นี้ทำหน้าที่เป็นเมล์เซิร์ฟเวอร์หรือผู้ให้บริการอีเมล์ ซึ่งจะมีกระบวนการที่ทำหน้าที่เป็นตัวแทนขนส่งเมล์ ทำงานอยู่บนทั้ง 2 ด้าน และรับส่งข้อมูลระหว่างกัน

โดยใช้ SMTP เมื่อได้รับเมล์มาแล้วก็จะเก็บเมล์เหล่านั้นไว้ในไดเร็คทอรี่ที่เป็นกล่องหรือตู้ไปรษณีย์ในเครื่องนั้น และรอจนกว่าผู้ใช้มาเปิดอ่าน ซึ่งมีได้ 3 วิธีด้วยกันคือ

- ผู้ใช้มีบัญชี (Account) บนเครื่องเมล์เซิร์ฟเวอร์ก็สามารถเปิดอ่านได้โดยใช้คำสั่งต่าง ๆ ของ Linux/Unix เช่น mail, pine และเมล์ที่ถูกอ่านจะถูกย้ายไปเก็บไว้ในเมล์บ๊อกซ์ของผู้ใช้แทนเมล์บ๊อกซ์ของระบบได้
- ผู้ใช้อยู่บนเครื่องไคล์แอนท์จะต้องโหลดเมล์ไปไว้ในเครื่องของตัวเองก่อนแล้วจึงเปิดอ่านได้
- ผู้ใช้รับส่งเมล์ผ่านตัวกลางที่เป็นเว็บเซิร์ฟเวอร์ซึ่งเมล์จะยังคงถูกเก็บไว้ที่เครื่องเมล์เซิร์ฟเวอร์

การทำงานของ SMTP จะทำหน้าที่ในการกำหนดว่าตัวแทนขนส่งเมล์แต่ละตัวจะติดต่อกันได้อย่างไรผ่านทางทีซีพี/ไอพี เมล์ที่ส่งไปนั้นอาจจะส่งตรงไปยังตัวแทนขนส่งเมล์ปลายทางเลย หรือว่าผ่านตัวแทนขนส่งเมล์หลายเครื่อง (หมายถึงผ่านรีเลย์โฮสต์หลายเครื่อง) โดยผ่านกระบวนการเก็บและส่งต่อ (Store and Forward) ก็ได้เช่นกัน

โปรโตคอล SMTP จะไม่สนใจว่าข้อความในเมล์เป็นอะไร แต่จำกัดว่า SMTP สามารถส่งได้แต่ข้อมูลที่เป็นข้อความแอสกี (ASCII) เท่านั้น ไม่สามารถส่งไฟล์ที่เป็นเพลง, หนัง, รูปภาพหรืออื่น ๆ ได้ ซึ่งถ้าเราต้องการส่งไฟล์เหล่านั้นผ่านทาง SMTP จะต้องแปลงไฟล์เหล่านั้นให้อยู่ในรูปของข้อความเสียก่อนและเมื่อส่งไปถึงปลายทางแล้วค่อยทำการแปลงกลับอีกที

นอกจากการใช้ SMTP เพื่อรับส่งเมล์ระหว่างเมล์เซิร์ฟเวอร์ด้วยกันแล้ว ยังใช้ในขณะที่เป็นไคล์แอนท์ส่งเมล์ไปยังเครื่องที่เป็นเมล์เซิร์ฟเวอร์ด้วย

#### 3.1.4.2 POP

POP หรือ Post Office Protocol คือโปรโตคอลที่ออกแบบมาให้ใช้สำหรับการรับเมล์จากเครื่องที่เป็นเมล์เซิร์ฟเวอร์มายังเครื่องของผู้ใช้ โดยทางฝั่งเซิร์ฟเวอร์จะมีกระบวนการที่เป็น POP เซิร์ฟเวอร์ขณะที่ทางฝั่งผู้ใช้มี POP ไคล์แอนท์ซึ่งในบางโปรแกรมที่ผู้ใช้อ่านและเขียนเมล์นั้นจะมี POP ไคล์แอนท์ฝังอยู่ในตัวอยู่แล้วไม่ได้แยกออกมาเป็นโปรแกรมหนึ่ง เมื่อผู้ใช้เชื่อมต่อไปที่ POP เซิร์ฟเวอร์อีเมล์ที่อยู่บนเมล์เซิร์ฟเวอร์จะถูกส่งมาเก็บไว้ในเครื่องของผู้ใช้เลย ดังนั้นเมื่อผู้ใช้จัดการกับเมล์ เช่น ลบเมล์หรือส่งต่อเมล์ก็จะทำกับเมล์ที่อยู่บนเครื่องของผู้ใช้เอง ส่วนเมล์บนเมล์เซิร์ฟเวอร์จะถูกลบทิ้งไปเมื่อมีการส่งให้ผู้ใช้เรียบร้อยแล้ว เว้นเสียแต่ว่าได้กำหนดเพิ่มเติมไว้ที่โปรแกรมเมล์ไคล์แอนท์ว่าอย่าให้ลบเมล์ออกจากเซิร์ฟเวอร์(Leave a copy of message on the server)

ในปัจจุบันโปรโตคอลมีออกมาหลายเวอร์ชัน แต่ที่นิยมกันคือ POP3 ซึ่งก็ยังมีข้อจำกัดในการใช้ คือขณะรับและส่งอีเมล์ ฝั่งผู้ใช้จะส่งรหัสผ่านของผู้ใช้ในรูปของข้อความไป ทำให้ไม่ปลอดภัย

นักหากมีการแอบดักข้อมูล เพราะฉะนั้น ตอนเซ็ต POP ไคล์แอนท์ เช่น MS outlook หรือโปรแกรมอื่น ๆ ควรจะเลือกใช้งานการเข้าใช้โดยใช้การตรวจสอบรหัสผ่าน (Log on using Secure Password Authentication: SPA) ด้วย แต่ต้องให้เมล์เซิร์ฟเวอร์มีการสนับสนุนการใช้ SPA ถึงจะใช้งานได้

**3.1.4.3 IMAP**

IMAP หรือ Internet Message Access Protocol เป็นโปรโตคอลที่เกิดหลัง POP เพื่อแก้ไขข้อจำกัดที่เกิดจาก POP นั่นเอง ทั้งนี้เพราะ POP จะใช้วิธีการโหลดเมล์ที่อยู่บนเซิร์ฟเวอร์มาเก็บไว้ยังเครื่องคอมพิวเตอร์ส่วนบุคคลของผู้ใช้แล้วลบเมล์นั้นทิ้ง (แต่ปัจจุบัน POP พัฒนาขึ้น คือสามารถกำหนดที่เมล์ไคล์แอนท์ได้ว่าจะให้ลบเมล์ทิ้งหรือไม่) ทำให้ผู้ใช้นั้นไม่สามารถอ่านเมล์จากเครื่องคอมพิวเตอร์ส่วนบุคคลเครื่องอื่น ๆ ได้อีก ต้องใช้เครื่องเดิมตลอดซึ่งเป็นปัญหาสำหรับผู้ใช้ที่มีเครื่องเครื่องคอมพิวเตอร์ส่วนบุคคลทั้งที่บ้านและที่ทำงาน หรือองค์กรที่มีเครื่องให้กับพนักงานไม่ครบทุกคน

การทำงานของ IMAP นั้นจะจัดการเมล์ที่อยู่บนเซิร์ฟเวอร์ เช่น อ่านหรือเขียนเมล์ ซึ่งเมล์เหล่านั้นจะยังคงอยู่บนเซิร์ฟเวอร์ ทำให้ผู้ใช้จะใช้เครื่องคอมพิวเตอร์ส่วนบุคคลเครื่องใดอ่านเมล์ก็ได้ หรือสั่งดาวน์โหลดเมล์ที่ต้องการมาเก็บในเครื่องพีซีของตนเองเหมือนกับการทำงานของ POP นอกจากนี้ยังสามารถกำหนดเมล์บ๊อกซ์หนึ่ง ๆ ให้กับผู้ใช้หลาย ๆ คนได้ โดยที่ผู้ใช้เหล่านั้นสามารถเปิดเมล์บ๊อกซ์อ่านได้พร้อม ๆ กัน สำหรับในกรณีเว็บเมล์เครื่องที่เป็นเว็บเซิร์ฟเวอร์ก็จะมีการติดต่อกับเมล์เซิร์ฟเวอร์โดยผ่านโปรโตคอล IMAP เช่นกัน

**3.1.5 คำนิยามของอีเมล์ขยะ**

G. Hulten จาก Msn Safety Team และ J. Godman จาก Microsoft Research [2] ได้แสดงตัวเลขของการสำรวจคำนิยามของอีเมล์ขยะพบว่า

- 92% เชื่อว่าอีเมล์ขยะคืออีเมล์ที่ประกอบไปด้วยเนื้อหาที่เกี่ยวข้องกับสื่อลามกอนาจาร
- 89% เชื่อว่าอีเมล์ขยะคืออีเมล์ที่ประกอบไปด้วยเนื้อหาที่เกี่ยวข้องกับข้อเสนอทางการเงินต่างๆ
- 76% เชื่อว่าอีเมล์ขยะคืออีเมล์ที่ประกอบไปด้วยเนื้อหาที่เกี่ยวข้องกับการเมืองหรือศาสนา
- 32% จะพิจารณาว่าอีเมล์เหล่านั้นเป็นอีเมล์ขยะถ้าหากมันมาจากผู้ส่งซึ่งทำธุรกิจอยู่ ส่วนคำนิยามซึ่งเป็นลักษณะทางกฎหมายกล่าวว่า อีเมล์ขยะคืออีเมล์ทางการค้าที่เราไม่ได้ให้ความสนใจจากใครก็ตามที่ไม่มีความสัมพันธ์ทางธุรกิจกันมาก่อน และ [2] ได้ให้คำนิยามว่า อีเมล์ขยะหมายถึง อีเมล์ใดๆ ก็ตามที่ผู้ใช้ในระบบรายงานให้ทราบ โดยผู้ใช้เหล่านั้นต้องได้รับการทดสอบอย่างใดอย่างหนึ่งแล้วว่าไม่ใช่สแปมเมอร์

## 3.2 องค์ประกอบของอีเมล์

อีเมล์ประกอบด้วย ส่วนหัวเรื่อง (Header) และส่วนเนื้อเรื่อง (Body) ดังรูปที่ 3.4

**Received:** from bright.net (unknown [83.134.197.119])
by zak.futurequest.net ([69.5.6.152])
with SMTP via TCP; 01 May 2005 14:44:44 -0000
**From:** <curluq@bright.net>
**To:** <rait@bruce-guenter.dyndns.org>
**Subject:** 3.25% intere/st ravj
**Date:** Sun, 1 May 2005 10:05:32 -0500
**Mime-Version:** 1.0
**Content-Type:** text/plain; charset=us-ascii
**Content-Length:** 463
**Lines:** 29

Hello
I have tried contacting you before in regards to the promised low interest mortga(ge rate. The good news is that you did qualify for the loan, for a rate of 4.2% ! You could receive over u380,000 for as little as i500 per month. Ba(d credit? Doesn't matter, low rates are fixed no matter what! To get a fr!ee, no obli,gation consultation click here http://goodwayz.com/x/loan.php?id=somb

Have a great day!
Ellen Horvan

**รูปที่ 3.4** แสดงตัวอย่างของอีเมล์ขยะ

### 3.2.1 ส่วนหัวเรื่อง

จากรูปเราพบว่าส่วนหัวเรื่องประกอบไปด้วยส่วนต่างๆ ดังนี้

1. ส่วน “ได้รับ” (Received) ในส่วนนี้ประกอบไปด้วย รายละเอียดของผู้ส่ง วิธีการที่ใช้ส่ง
2. ส่วน “จาก” (From) ส่วนนี้จะแสดงที่อยู่อีเมล์ของผู้ส่ง
3. ส่วน“ถึง” (To) ส่วนนี้จะแสดงให้เห็นว่า ผู้ส่งได้ทำการส่งข้อความนี้ไปถึงใครบ้าง
4. ส่วน “วันที่” (Date) แสดงวันที่ที่ส่ง
5. ส่วน “ชื่อเรื่อง” (Subject) ส่วนชื่อเรื่องคือส่วนที่จะแสดงให้เห็นเป็นส่วนแรกก่อนที่ผู้ใช้คลิกเข้าไปจึงจะพบกับส่วนเนื้อหาซึ่งเป็นใจความหลักของอีเมล์ ชื่อเรื่องจึงเป็นเหมือนส่วนที่ใช้

สำหรับแสดงเนื้อหาโดยสรุปของอีเมล์นั่นเอง มีผู้นำส่วนชื่อเรื่องมาเป็นส่วนร่วมในการวิเคราะห์ว่าอีเมล์เป็นอีเมล์ขยะหรืออีเมล์ดีเป็นจำนวนมาก เนื่องจากชื่อเรื่องก็อาจสื่อความหมายไปถึงเนื้อความในอีเมล์ได้เช่นกัน

6. ส่วน “รหัสข้อความ” (Message ID) แสดงรหัสของข้อความหรืออีเมล์นั้นๆ

7. ส่วน “เวอร์ชันเอ็มไอเอ็มอี” (MIME-Version) ชื่อเต็มว่าMultipurpose Internet Mail Extensions เป็นมาตราฐานของรูปแบบของอีเมล์

8. ส่วน “ชนิดเนื้อหา” (Content-Type) เป็นส่วนของข้อมูลที่บอกว่าข้อความนี้ถูกแสดงให้เห็นด้วยวิธีใด โดยทั่วๆ ไปแล้วจะเป็นชนิดเอ็มไอเอ็มอี

ในส่วนหัวเรื่องนี้มักถูกนำไปใช้ในการวิเคราะห์อีเมล์ขยะทางฝั่งเซิร์ฟเวอร์เป็นส่วนใหญ่ เนื่องจากค่อนข้างยุ่งยากและไม่คุ้มหากจะใช้เป็นการส่วนตัวหรือฝั่งลูกข่าย แต่สำหรับการวิเคราะห์ที่ต้องการความถูกต้องและปลอดภัยสูง การกรองที่ฝั่งเซิร์ฟเวอร์จึงต้องเข้ามงวดกว่าจึงจำเป็นต้องรับรู้ข้อมูลที่บ่งบอกถึงที่มาที่ไปของอีเมล์เพิ่มเติมในการพิจารณาด้วย

### 3.2.2 ส่วนเนื้อเรื่อง

ส่วนเนื้อเรื่อง คือส่วนของข้อความที่ผู้ใช้งานอีเมล์ใช้ติดต่อสื่อสารกัน และเป็นส่วนที่นักวิจัยมักนำมาใช้พิจารณาเนื่องจากง่ายต่อตัวกรองในการเรียนรู้ เนื่องด้วยเหตุนี้อีเมล์บางส่วนจึงจำเป็นต้องทำให้เนื้อหาของอีเมล์นั้นคลุมเครือเพื่อหลบหนีบรรดาตัวกรองซึ่งมีอยู่มากมาย[2] ได้จำแนกวิธีต่างๆ ที่ใช้ทำให้อีเมล์คลุมเครือ เช่นการทำส่วนเนื้อหาของอีเมล์ให้เป็นรูปภาพ (Content in Images) การเพิ่มคำดีดีที่ไม่ปะติดปะต่อเข้าไปในอีเมล์ (Good Word Chaff) สุ่มเพิ่มประโยคที่นำมาจากเมล์ดี (Content Chaff) เพราะฉะนั้นในบางครั้งการกรองอีเมล์โดยพิจารณาจากตัวอักษรอย่างเดียวอาจไม่พอสำหรับอีเมล์ขยะที่มีหน้าตาแปลกดังที่ได้กล่าวข้างต้น

## 3.3 ผลกระทบของอีเมล์ขยะ

ผลกระทบต่อของอีเมล์ขยะนั้นมีหลายอย่างด้วยกัน โดยที่เห็นได้ชัดมีดังต่อไปนี้

1. สิ้นเปลืองเวลาของผู้ใช้ในการจัดการคัดแยกและลบอีเมล์ขยะทิ้ง ผลของอีเมล์ขยะอาจดูไม่ได้ร้ายแรงอะไรถ้าคุณเป็นผู้ใช้ตามบ้าน เพียงแค่ถ้ามีอีเมล์ขยะมาคุณก็แค่ลบเมล์นั้นทิ้งก็ไม่มีอะไรเกิดขึ้น แต่ถ้าเป็นองค์กรธุรกิจ การที่พนักงานต้องคอยมานั่งลบอีเมล์ขยะคงไม่ใช่เรื่องที่น่ายินดีนัก มีผลทำให้การทำงานขององค์กรล่าช้า และส่งผลเสียแก่องค์กรไม่น้อย ตัวอย่างเช่น ถ้าบริษัทหนึ่งให้บริการออนไลน์ แล้วมีการส่งอีเมล์ขยะมาสักวันละ 100,000 ฉบับ ซึ่งอาจทำให้เมล์บ๊อกซ์ของบริษัทเต็มส่งผลให้พนักงานแต่ละคนต้องเสียเวลาในการลบเมล์ขยะเมล์ละ 2 วินาที ซึ่งเมื่อมาคิดแล้วเวลารวมที่พนักงานในบริษัทต้องเสียไปกับการลบอีเมล์ขยะถึง 555 ชั่วโมงในการลบหมด

2. สิ้นเปลืองแบนด์วิธทำให้เมล์เซิร์ฟเวอร์ต้องเสียแบนด์วิธไปเป็นจำนวนมากพอๆกับขนาดของอีเมล์ขยะนั้น หากแบนด์วิธมีอยู่อย่างจำกัดก็จะส่งให้อีเมล์อื่นที่เข้ามาจะต้องใช้เวลานานกว่าปกติหรืออาจจะไม่ได้รับในที่สุด

3. สิ้นเปลืองการประมวลผลของหน่วยประมวลผลกลางที่เมล์เซิร์ฟเวอร์

4. สิ้นเปลืองเนื้อที่ในเมล์บ็อกซ์ ซึ่งจะมีผลกระทบมากหากเมล์บ็อกซ์มีการจำกัดเนื้อที่ของผู้ใช้ หากผู้ใช้ทิ้งไว้ไม่ได้มาตรวจสอบบ่อยๆก็จะทำให้เนื้อที่หมดไปได้ หากเป็นอีเมล์ทางธุรกิจที่สำคัญก็จะทำให้เสียหายต่อธุรกิจได้

5. ส่งผลต่อผู้ให้บริการเซิร์ฟเวอร์ที่มีการตั้งค่าในการรีเลย์ไว้ไม่จำกัดกลุ่มที่แน่นอน คือจะทำให้ สแปมเมอร์สามารถใช้เซิร์ฟเวอร์นั้นทำการส่งอีเมล์ขยะออกไป ซึ่งเมื่อมีการสืบต้นตอของตัวแทนขนส่งเมล์ (MTA) ก็จะทำให้เซิร์ฟเวอร์นั้นถูกบล็อกทำให้ไม่สามารถส่งอีเมล์ได้

## 3.4 ความรู้พื้นฐานเกี่ยวกับทฤษฎีเบย์เซียน (Bayesian Theorem)

ในอดีตอีเมล์ขยะยังไม่ค่อยจะมีผลกระทบต่อระบบคอมพิวเตอร์มากนัก การกรองอีเมล์ขยะในอดีตจึงมีการใช้กฎเกณฑ์อย่างง่ายๆ ในการจำแนก วิธีการที่ใช้เช่น การตั้งกฎในการจำแนกรูปแบบของอีเมล์ขยะ (Pattern-Matching) โดยมีการตั้งกฎที่ระบุว่าอีเมล์ลักษณะไหนที่เป็นอีเมล์ขยะ แต่เนื่องจากในปัจจุบันผู้ส่งอีเมล์ขยะ สามารถส่งอีเมล์ขยะที่มีความสามารถในการหลบเลี่ยงตัวกรอง จึงทำให้วิธีการจำแนกรูปแบบของอีเมล์ขยะเริ่มที่จะไม่มีประสิทธิภาพ เพราะวิธีการนี้เป็นวิธีที่มีการตั้งกฎเกณฑ์ที่ตายตัว ซึ่งเมื่อผู้ส่งอีเมล์ขยะเปลี่ยนแปลงรูปแบบในการส่งก็จำเป็นต้องเพิ่มกฎใหม่เข้าไปเรื่อยๆ ทำให้ยากต่อการเปลี่ยนแปลงตามความสามารถในการหลบเลี่ยงที่เปลี่ยนไปของผู้ส่งอีเมล์ขยะ จึงมีการคิดค้นวิธีการในการกรองอีเมล์ขยะที่มีความสามารถเพิ่มขึ้น โดยมีนำทฤษฎีต่างๆ เข้ามาเพื่อใช้ในการพัฒนาตัวกรองอีเมล์ขยะเพื่อให้มีความสามารถที่เพิ่มขึ้น

หนึ่งในวิธีที่ใช้ในปัจจุบันและมีประสิทธิภาพคือการนำการวิเคราะห์ทางสถิติเข้ามามีส่วนช่วยในการเรียนรู้ของอีเมล์ ซึ่งเป็นวิธีการที่มีประสิทธิภาพในการจำแนกประเภทของอีเมล์ขยะคือการจำแนกโดยใช้ทฤษฎีเบย์เซียน ตัวกรองอีเมล์ขยะที่พัฒนาขึ้นโดยการใช้ทฤษฎีเบย์เซียนจะมีพื้นฐานของการนำความน่าจะเป็นของอีเมล์มาคำนวณ ซึ่งเป็นที่วิธีการในการคัดเลือกคีย์เวิร์ดเพื่อนำมาใช้ในการคำนวณ โดยการคัดเลือกคีย์เวิร์ดจะขึ้นกับหลักการในการจำแนกคำ รวมทั้งการตั้งกฎของการจำแนกคำโดยผู้พัฒนาเป็นผู้จัดการ การใช้หลักการทางสถิติมาใช้ในการสร้างตัวกรองเป็นสิ่งที่ค่อนข้างสมเหตุสมผล เพราะว่าความแตกต่างกันระหว่างอีเมล์ทั่วไปกับอีเมล์ขยะ ค่อนข้างจะแยกออกจากกันได้ยาก ซึ่งอีเมล์ขยะทั่วไปจะมีรูปแบบหรือว่าหลักการที่คล้ายคลึงกันกับอีเมล์ทั่วไป ซึ่งสุดท้ายต้องขึ้นกับผู้ใช้เป็นผู้ระบุเองว่าอีเมล์ฉบับไหนที่เป็นอีเมล์ขยะ หรือว่าอีเมล์ทั่วๆไป และผู้ใช้แต่ละคนอาจจะระบุได้แตกต่างกัน ซึ่งอีเมล์ฉบับหนึ่งเป็นอีเมล์ขยะของผู้ใช้คนหนึ่ง แต่ในขณะที่ผู้ใช้อีกคนอาจจะบอกว่าเป็นอีเมล์ปกติ

จากการที่ผู้ใช้เป็นผู้กำหนดนิยามของอีเมล์ขยะเองทำให้สังเกตได้ว่าสิ่งหนึ่งที่แตกต่างกันระหว่างอีเมล์ทั่วไปกับอีเมล์ขยะก็คือ เนื้อหาที่อยู่ภายในอีเมล์ วิธีการทางด้านสถิติจะมีข้อเสียก็คือการใช้งานตัวกรองที่สร้างขึ้นจะต้องให้อีเมล์มีการเรียนรู้ในรูปแบบต่างๆ ซึ่งต้องใช้ระยะเวลาหนึ่งการสอนอีเมล์ให้แก่ระบบตัวกรอง ซึ่งทฤษฎีเบย์เซียนจะมีรายละเอียดดังต่อไปนี้

**3.4.1 ทฤษฎีเบย์เซียน**

ทฤษฎีเบย์จะเกี่ยวข้องกับเงื่อนไขและสถิติของเหตุการณ์ A และ B ดังสมการที่ (3.1)

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)} \tag{3.1}$$

เมื่อ

$\Pr(A)$ คือความน่าจะเป็นที่มาก่อน (Prior Probability) หรือค่าความน่าจะเป็นซึ่งมีความสำคัญน้อย (Marginal Probability) ของ A

$\Pr(A \mid B)$ คือความน่าจะเป็นแบบมีเงื่อนไข (Conditional Probability) ของ A เมื่อให้ B หรือเรียกว่าความน่าจะเป็นที่มาที่หลัง (Posterior Probability) เนื่องจากถูกแปลงหรือเป็นอิสระจากค่าของ B ที่ได้กำหนดไว้

$\Pr(B \mid A)$ คือความน่าจะเป็นแบบมีเงื่อนไขของ B เมื่อให้ A

$\Pr(B)$ คือความน่าจะเป็นที่มาก่อน (Prior Probability) หรือค่าความน่าจะเป็นซึ่งมีความสำคัญน้อย (Marginal Probability) ของ B ซึ่งกระทำตัวเป็นนอร์มัลไลซิ่งคอนสแตนท์ (Normalizing Constant) ถอดความทฤษฎีใหม่ได้ดังสมการที่ (3.2)

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Normalizing Constant}} \tag{3.2}$$

นั่นคือความน่าจะเป็นที่เกิดทีหลังคือสัดส่วนของจำนวนครั้งที่คล้ายกับความน่าจะเป็นที่เกิดขึ้นก่อน ในทางตรงกันข้าม อัตราส่วน $P(B \mid A) / P(B)$ ในบางครั้งเราจะเรียกว่าความคล้ายมาตราฐาน (Standardized Likelihood) ดังนั้นทฤษฎีนี้จึงถูกถอดความใหม่ได้ดังสมการที่ (3.3)

$$\text{Posterior} = \text{Standardized Likelihood} \times \text{Prior} \tag{3.3}$$

จากนั้นเราจะดีไรฟ์ (Derive) ทฤษฎีโดยเริ่มจากนิยามของความน่าจะเป็นแบบมีเงื่อนไข ความน่าจะเป็นของเหตุการณ์ A เมื่อให้เหตุการณ์ B เป็นไปตามสมการที่ (3.4) คือ

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} \tag{3.4}$$

ในทำนองเดียวกัน ความน่าจะเป็นของเหตุการณ์ B เมื่อให้เหตุการณ์ A เป็นไปตามสมการที่ 3.5 คือ

$$\mathrm{P}(B \mid A) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(A)} \tag{3.5}$$

เมื่อนำ 2 สมการนี้มาจัดเรียงใหม่จะเป็นไปตามสมการที่ (3.6) คือ

$$\mathrm{P}(A \mid B)P(B) = \mathrm{P}(A \cap B) = P(B \mid A)\,\mathrm{P}(A) \tag{3.6}$$

หารทั้งสองข้างด้วย $\Pr(B)$ จะได้ทฤษฎีของเบย์จะเป็นไปตามสมการที่ (3.7) คือ

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(B \mid A)\,\mathrm{P}(A)}{\mathrm{P}(B)} \tag{3.7}$$

**3.4.2 การนำทฤษฎีเบย์เซียนมาใช้ในการกรองอีเมล์ขยะ**

ในการนำทฤษฎีของเบย์มาประยุกต์ใช้กับตัวกรองอีเมล์ขยะนั้น นิยมใช้ทฤษฎีเบย์อย่างง่าย (Naïve Bayes)[2] ซึ่งบางครั้งเราอาจเรียกว่าเบย์เซียนเทคนิค โดยทฤษฎีเบย์อย่างง่ายนั้นจะมองว่าคำแต่ละคำเป็นอิสระจากกัน

ถ้าต้องการ P(spam|mail) ซึ่งหมายถึงความน่าจะเป็นที่อีเมล์จะเป็นขยะ สามารถหาได้โดยใช้กฎของเบย์ซึ่งคำนวณตามสมการที่ 3.8 โดยที่ค่า P(ham) หรือความน่าจะเป็นของอีเมล์ดี, P(spam) หรือความน่าจะเป็นของอีเมล์ขยะ, P(mail) หรือความน่าจะเป็นของอีเมล์ หาได้จาก สมการที่ (3.9)-(3.11) ตามลำดับ

$$P(spam \mid mail) = \frac{P(mail \mid spam) \times P(spam)}{P(mail)} \tag{3.8}$$

$$P(ham) = \frac{nMailHam}{nMailTotal} \tag{3.9}$$

$$P(spam) = \frac{nMailSpam}{nMailTotal} \tag{3.10}$$

$$P(mail) = P(mail \mid spam) \times P(spam) + P(mail \mid ham) \times P(ham) \qquad (3.11)$$

จากสมมติฐานความเป็นอิสระที่กล่าวว่า ความน่าจะเป็นของแต่ละตัวเป็นอิสระจากกัน (ข้อสมมติฐานที่ผิด) ทำให้ได้สมการที่ (3.12) และ (3.13)

$$P(mail \mid spam) \approx P(word_1 \mid spam) \times P(word_2 \mid spam) \times ... \times P(word_n \mid spam) \qquad (3.12)$$

$$P(mail \mid ham) \approx P(word_1 \mid ham) \times P(word_2 \mid ham) \times ... \times P(word_n \mid ham) \qquad (3.13)$$

อีเมล์ประกอบด้วยคำหลายๆคำ(Words) เมื่อต้องการทราบว่าอีเมล์ฉบับใดเป็นอีเมล์ขยะ จะพิจารณาจากความน่าจะเป็นที่จะเป็นอีเมล์ขยะของคำแต่ละคำในอีเมล์แล้วจึงหาความน่าจะเป็นรวมของอีเมล์ทั้งฉบับโดยการนำความน่าจะเป็นของคำขยะที่ได้มาคูณกัน เช่นเดียวกัน ถ้าต้องการทราบว่าอีเมล์ฉบับใดเป็นอีเมล์ดีให้พิจารณาจากความน่าจะเป็นที่จะเป็นอีเมล์ดีของคำดีในอีเมล์แล้วทำเช่นเดียวกันกับการหาความน่าจะเป็นของอีเมล์ขยะ

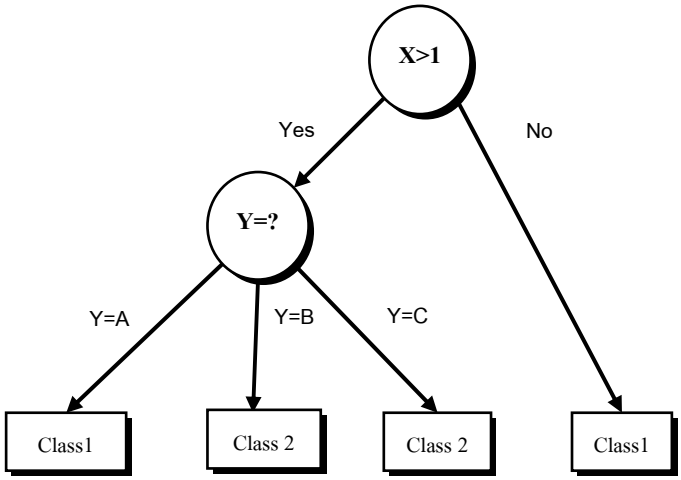
เมื่อได้ความน่าจะเป็นจะเป็นของอีเมล์ที่จะเป็นอีเมล์ขยะ $P(mail \mid spam)$ และความน่าจะเป็นของอีเมล์ที่จะเป็นอีเมล์ดี $P(mail \mid ham)$ ให้นำมาเปรียบเทียบกันถ้าความน่าจะเป็นของฝั่งอีเมล์ขยะมีค่ามากกว่าให้ตอบว่าเป็นอีเมล์ที่นำมาทดสอบเป็นอีเมล์ขยะ ถ้าไม่ก็ตอบว่าเป็นอีเมล์ดี

## 3.5 ความรู้พื้นฐานเกี่ยวกับแขนงการตัดสินใจ (Decision Tree)

แขนงการตัดสินใจคือโมเดลสำหรับการทำนายซึ่งใช้ Top-Down Strategy มาใช้ในการค้นหาคำตอบในเส้นทางของคำตอบที่สารถเป็นไปได้ องค์ประกอบของแขนงการตัดสินใจประกอบด้วย

1. โหนด (Node) คือสิ่งที่ถูกทดสอบ หรือเงื่อนไข
2. แขนงหรือกิ่ง (Branches) คือ ผลลัพธ์ที่เป็นไปได้ทั้งหมดของโหนดนั้น

ตัวอย่างของแขนงการตัดสินใจแสดงดังรูปที่ 3.5




**รูปที่ 3.5** แสดงตัวอย่างของแขนงการตัดสินใจ

แขนงการตัดสินใจสามารถนำมาประยุกต์เป็นโครงสร้างของกฎได้ เช่น

IF (Condition) THEN (Action)

IF (Condition1) AND (Condition2)THEN (Action)

## 3.6 ความรู้พื้นฐานเกี่ยวกับเจเนติกอัลกอริทึม

ในปี ค.ศ. 1975 จอห์น ฮอลแลนด์และคณะ [15] จากมหาวิทยาลัยมิชิแกนได้ตีพิมพ์หนังสือการปรับตัวในสิ่งแวดล้อมและระบบเทียม (Adaptation in Natural and Artificial System) ขึ้น ซึ่งเป็นครั้งแรกที่นำแนวความคิดของการเลียนแบบวิวัฒนาการธรรมชาติมากมายมารวมไว้ด้วยกันบนคอมพิวเตอร์ ซึ่งเตรียมสิ่งสำคัญต่าง ๆ ในการวิเคราะห์ปัญหาทางคณิตศาสตร์ เสมือนกับการเลียนแบบการวิวัฒนาการบนคอมพิวเตอร์ และเมื่อไม่นานมานี้งานมากมายที่ได้มาจากการเลียนแบบการวิวัฒนาการบนคอมพิวเตอร์ก็ถูกสร้าง จุดประสงค์เดียวเพื่อความเข้าใจในพันธุ์กรรม ( Genetic ) และวิวัฒนาการ ( Evolution ) หนังสือของฮอลแลนด์แสดง เคร้าโครงว่ากระบวนการสามารถใช้ในการแก้ปัญหาในโลกที่แท้จริงได้อย่างไรด้วยเทคนิคของการวิวัฒนาการ ฮอลแลนด์ได้ให้ความหมายของเจเนติกอัลกอริทึมว่า เป็นอัลกอริทึมสำหรับการค้นหาข้อมูลและการค้นหาคำตอบที่ดีที่สุด (Optimization) ซึ่งได้รับแนวคิดมาจากกลไกการคัดเลือกสายพันธุ์ตามธรรมชาติ (Natural Selection) และธรรมชาติทางพันธุกรรม (Natural Genetic) คือสิ่งมีชีวิตใดที่มีความแข็งแรงกว่าย่อมมีโอกาสอยู่รอดได้มากกว่านั่นหมายถึงการมีสายพันธุ์ที่ดีย่อมมีโอกาสจะได้รับกา

คัดเลือกนำมาเป็นต้นแบบ เพื่อถ่ายทอดลักษณะดี ๆ ของสายพันธุ์เหล่านั้นไปยังรุ่นต่อๆ ไป มากกว่า มีโอกาสในการอยู่รอดสูง ส่วนสายพันธุ์ที่ไม่ดีก็จะไม่ได้รับการคัดเลือก หรือได้รับการคัดเลือกน้อยกว่า และจะค่อยๆ สูญพันธุ์ไปในที่สุดเจเนติกอัลกอริทึมได้นำกระบวนการวิวัฒนาการของสิ่งมีชีวิตมาประยุกต์ใช้กับงานด้านปัญญาประดิษฐ์ เพื่อค้นหาคำตอบของปัญหาต่าง ๆ โดยเจเนติกอัลกอริทึมเป็นรูปแบบของเทคนิคการค้นหาซึ่งใช้ในการค้นหาคำตอบจากจำนวนคำตอบที่เป็นไปได้ทั้งหมดของการแก้ปัญหานั้นๆ เพื่อให้ได้คำตอบที่เหมาะสมกับปัญหาโดยอาศัยข้อมูลในการช่วยค้นหา ซึ่งข้อมูลหรือวิธีการที่ใช้นี้จำลองมาจากกฎเกณฑ์การคัดเลือกสายพันธุ์ตามธรรมชาตินั่นเอง

เจเนติกอัลกอริทึมมีองค์ประกอบที่สำคัญ 5 องค์ประกอบ ได้แก่

1. นำเสนอปัญหาด้วยรูปแบบโครโมโซม และทางเลือกที่เป็นไปได้ของแต่ละปัญหา
2. วิธีการสร้างประชากรต้นกำเนิด (Initial population) ของทางเลือกที่เป็นไปได้
3. ฟังก์ชันความเหมาะสม (Fitness function) เพื่อให้คะแนนแต่ละทางเลือก
4. เจเนติกโอเปอเรเตอร์ (Genetic Operator) ซึ่งใช้ปรับเปลี่ยนองค์ประกอบของข้อมูลตลอดกระบวนการ
5. ค่าพารามิเตอร์ต่างๆ ซึ่งต้องใช้ในเจเนติกอัลกอริทึม เช่น ขนาดของประชากร ความน่าจะเป็นของการใช้เจเนติกโอเปอเรเตอร์ เป็นต้น

เจเนติกอัลกอริทึมแตกต่างจากวิธีการโดยทั่วไป คือ

1. เป็นวิธีการที่ค้นหาคำตอบภายใต้โครงสร้างของปัญหาอันเกิดจากการเข้ารหัส (Encoding) รูปแบบปัญหา
2. โครงสร้างจากกลุ่มตัวแปรต่างๆ ของปัญหานั้น ไม่ใช่การค้นหาคำตอบจากค่าของกลุ่มตัวแปรนั้นโดยตรง
3. ทำการค้นหาคำตอบจากกลุ่มประชากรคำตอบ(Population)แทนการหาคำตอบใดคำตอบหนึ่ง
4. ทำการค้นหาคำตอบจากผลลัพธ์ของกลุ่มค่าตัวแปรที่เป็นฟังก์ชันเป้าหมาย (Objective Function) ไม่สนใจข้อมูลข่าวสารแวดล้อมอื่นๆ
5. ใช้ความน่าจะเป็น (Probability) ในการค้นหาคำตอบ

### 3.6.1 พันธุศาสตร์ทางชีววิทยากับเจเนติกอัลกอริทึม

ตามธรรมชาติ สิ่งมีชีวิตแต่ละชนิดจะมีโครงสร้างและพฤติกรรมที่แตกต่างกัน อันเนื่องมาจากสภาพแวดล้อมการดำรงชีวิตที่แตกต่างกัน ลักษณะที่แตกต่างกันนี้มีผลต่ออัตราการมีชีวิตรอดและอัตราการสืบพันธุ์ โดยสิ่งมีชีวิตมีแนวโน้มจะถ่ายทอดคุณลักษณะพิเศษให้กับประชากรรุ่นลูกหลาน (Offspring) และให้กำเนิดสิ่งมีชีวิตที่มีลักษณะพิเศษแตกต่างไปจากเดิมที่มีคุณสมบัติ

เหมาะสม เพื่อให้สามารถดำรงอยู่รอดได้ต่อไปในสภาพแวดล้อมของสิ่งมีชีวิตนั้นๆ ประชากรจะมีแนวโน้มที่จะมีคุณลักษณะที่เหมาะสมต่อการดำรงชีวิตมากกว่ารุ่นบรรพบุรุษ เมื่อเวลาผ่านไปหลายๆ รุ่น (Generation) ของวิวัฒนาการ สิ่งมีชีวิตนั้นก็จะได้สายพันธุ์ใหม่ที่เหมาะสมกับสภาพแวดล้อมมากยิ่งขึ้น ตัวอย่างของวิวัฒนาการเหล่านี้ เช่น มีการสันนิษฐานว่ายีราฟในสมัยโบราณอาจจะมีลำคอไม่ยาวเท่ากับยีราฟในยุคปัจจุบัน ยีราฟดำรงชีวิตด้วยการกินใบไม้ตามยอดไม้ เมื่อจำนวนประชากรยีราฟมีมากขึ้น การแก่งแย่งอาหารเพื่อความอยู่รอดจึงสูงขึ้นตาม ยีราฟตัวที่สามารถกินยอดไม้สูงๆ เท่านั้นจึงจะมีชีวิตอยู่ต่อไป คุณสมบัติที่ดีในการดำรงชีวิตนี้จึงถูกคัดเลือกและถ่ายทอดมายังยีราฟรุ่นลูกหลาน นั่นหมายถึงยีราฟที่คอยาวเท่านั้นที่จะมีโอกาสหาอาหาร และรอดชีวิตสูงกว่ายีราฟคอสั้น

จากที่กล่าวมาข้างต้นว่าเจเนติกอัลกอริทึมเป็นวิธีการที่เลียนแบบมาจากหลักการทางชีววิทยา จึงมีการนำศัพท์ต่างๆ ในด้านพันธุศาสตร์มาประยุกต์ใช้ในกระบวนการเจเนติกอัลกอริทึม ดังต่อไปนี้

#### 3.6.1.1 พันธุศาสตร์ทางชีววิทยา

ในแต่ละเซลล์ (Cell) ของสิ่งมีชีวิตจะประกอบไปด้วยหน่วยย่อยที่มีความสำคัญมากอยู่ในนิวเคลียส (Nucleus) ของเซลล์ นั่นคือโครโมโซม (Chromosome) แต่ละโครโมโซมจะประกอบไปด้วยยีนส์ (Genes) ซึ่งเป็นหน่วยเก็บลักษณะต่างๆ ของสิ่งมีชีวิต ภายในยีนส์จะมีค่าแสดงลักษณะต่างๆ หรือแอลลี (Allele) โดยตำแหน่งของยีนตาละยีนในโครโมโซมจะเรียกว่าโลคัส (Locus) รูปแบบของยีนส์ที่แตกต่างกันเรียกว่าจีโนไทป์ (Genotype) ส่วนลักษณะภายนอกที่ปรากฏเรียกว่าฟีโนไทป์ (Phenotype) [16]

#### 3.6.1.2 พันธุศาสตร์ทางเจเนติกอัลกอริทึม

สำหรับเจเนติกอัลกอริทึม ตัวแปรของปัญหาจะถูกแปลงให้อยู่ในรูปของสตริง (String) ว่าโครโมโซม ภายในโครโมโซมจะประกอบไปด้วยอักขระ (Character) หรือบิท (Bit) แต่ละตำแหน่งของบิทจะเก็บค่าอักขระ (Character value) หรือค่าของบิท (Bit value) ที่แสดงโครงสร้างของแต่ละโครโมโซมของปัญหาที่แตกต่างกัน สรุปความหมายของคำสำคัญต่างๆ ได้ดังตารางที่ 1

**ตารางที่ 3.1** คำสำคัญที่ใช้ในกระบวนการเจเนติกอัลกอริทึม

| Natural Genetic Terms | Genetic Algorithm Terms |
|---|---|
| Chromosome | String |
| Gene | Feature, Character, Bit |
| Allele | Feature value, Character value, Bit value |
| Locus | String position |
| Genotype | Structure |
| Phenotype | Decoded structure, Alternative solution |

### 3.6.2 ขั้นตอนการทำงานของเจเนติกอัลกอริทึม

#### 3.6.2.1 การกำหนดรูปแบบโครโมโซม (Chromosome Representation)

การกำหนดปัญหาโดยใช้เจเนติกอัลกอริทึมนั้น จะต้องมีการนำปัญหามาเข้ารหัสข้อมูลให้อยู่ในรูปแบบโครโมโซมที่เหมาะสม เช่น อาจนำเสนอในรูปแบบของเลขฐานสอง เลขจำนวนจริง ตัวอักษร) การสลับลำดับกันในพีชคณิต และแบบทรี (Tree) เป็นต้น สำหรับเจเนติกอัลกอริทึมจะใช้กระบวนการเจเนติกอัลกอริทึมแบบง่าย (Simple Genetic Algorithm) [17] ได้แสดงตัวอย่างการเข้ารหัสแบบไบนารี การเข้ารหัสแบบสลับลำดับ (Permutation Encoding) การเข้ารหัสแบบค่า (Value Encoding) และการเข้ารหัสแบบทรี (Tree Encoding) ดังแสดงในรูปที่ 3.6-3.9 ตามลำดับ

| | |
|---|---|
| Chromosome A | 1011001011001010111100101 |
| Chromosome B | 1111111000001100000011111 |

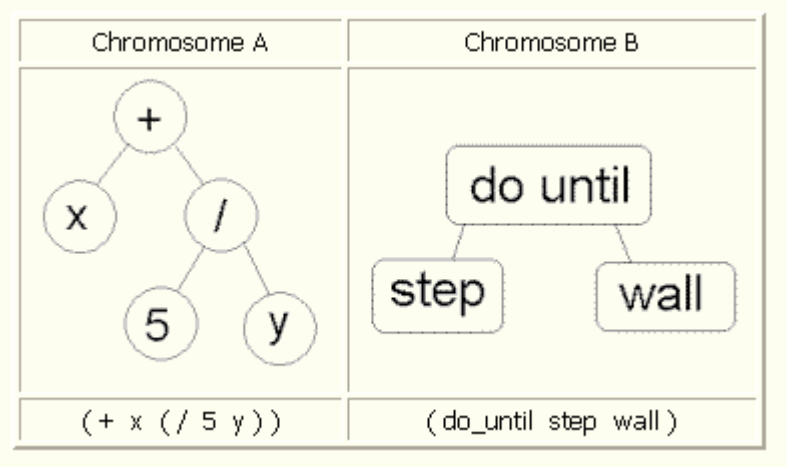
**รูปที่ 3.6** แสดงการเข้ารหัสแบบไบนารี

| | |
|---|---|
| Chromosome A | 1 5 3 2 6 4 7 9 8 |
| Chromosome B | 8 5 6 7 2 3 1 4 9 |

**รูปที่ 3.7** แสดงการเข้ารหัสแบบสลับลำดับ

| | |
|---|---|
| Chromosome A | 1.2324 5.3243 0.4556 2.3293 2.4545 |
| Chromosome B | ABDJEIFJDHDIERJFDLDFLFEGT |
| Chromosome C | (back), (back), (right), (forward), (left) |

**รูปที่ 3.8** แสดงการเข้ารหัสแบบค่า




**รูปที่ 3.9** แสดงการเข้ารหัสแบบทรี

**3.6.2.2 ประชากร (Population)**

ประชากรในกระบวนการเจเนติกอัลกอริทึมจะแบ่งออกเป็น 2 กลุ่มคือ ประชากรรุ่นเก่า (Old Population) และประชากรรุ่นใหม่ (New Population) ประชากรรุ่นเก่าจะถูกสร้างขึ้นมาเพื่อที่จะคัดเลือกไปเป็นประชากรรุ่นใหม่ และสำหรับประชากรต้นกำเนิด (Initial Population) ซึ่งเป็นประชากรรุ่นแรกในกระบวนการสามารถทำได้โดยการสุ่มสร้างค่าที่เป็นไปได้ของแต่ละบิทของแต่ละโครโมโซมตามที่ต้องการ

**3.6.2.3 กำหนดฟังก์ชันความเหมาะสม (Fitness Function)**

การกำหนดฟังก์ชันความเหมาะสมคือ การสร้างฟังก์ชันเพื่อคำนวณหาความเหมาะสมของประชากรว่าเหมาะสมที่จะถูกคัดเลือกมาเพื่อสร้างประชากรรุ่นต่อไปมากน้อยเพียงใด อาจเป็นการวัดจากค่าความเหมาะสมที่สูงสุด (Max) หรือเป็นค่าความเหมาะสมที่ต่ำสุด (Min) ก็ได้ โดยฟังก์ชันความเหมาะสมนั้นจะแตกต่างกันออกไปสำหรับแต่ละปัญหา

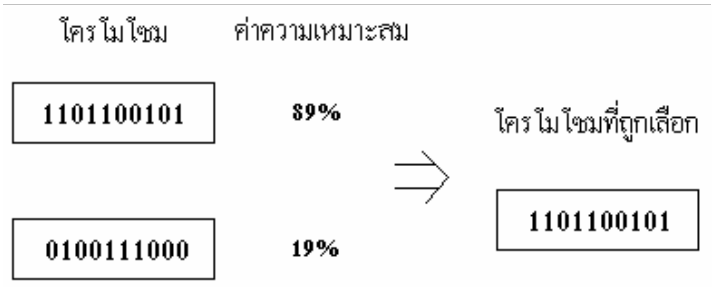
**3.6.2.4 การวิเคราะห์ค่าความเหมาะสม (Fitness Evaluation)**

เป็นขั้นตอนการถอดรหัสโครโมโซม โดยนำค่าที่ได้จากการถอดรหัสนี้ไปแทนค่าในฟังก์ชันความเหมาะสมของปัญหา ผลลัพธ์ที่ได้จากการคำนวณนี้เรียกว่า ค่าความเหมาะสม ซึ่งค่านี้จะเป็นสิ่งที่แสดงว่า แต่ละโครโมโซมมีความเหมาะสมที่จะนำมาใช้แก้ปัญหามากน้อยเพียงใด หรืออาจเปรียบเทียบได้กับค่าความสามารถในการอยู่รอดของแต่ละโครโมโซม และเป็นการกำหนดโอกาสหรือสัดส่วนที่แต่ละโครโมโซมจะถูกคัดเลือกมาเป็นต้นแบบในการให้กำเนิดประชากรรุ่นต่อไป

**3.6.2.5 การคัดเลือก (Selection)**

เป็นขั้นตอนที่จำลองแบบการคัดเลือกประชากรตามธรรมชาติ เพื่อคัดเลือกโครโมโซมรุ่นเก่าให้เป็นโครโมโซมต้นแบบหรือโครโมโซมพ่อแม่ เพื่อใช้ในการสร้างประชากรรุ่นลูกต่อไป

สำหรับการคัดเลือกทำได้โดยการวัดค่าความเหมาะสมของแต่ละโครโมโซมโดยวิธีใดวิธีหนึ่ง แล้วคัดเลือกโครโมโซมจำนวนหนึ่งที่มีค่าความเหมาะสมเป็นที่น่าพอใจเก็บไว้ ดังตัวอย่างในรูปที่ 3.10 ดยอาจจะคัดเลือกเอาเฉพาะโครโมโซมที่มีค่าความเหมาะสมสูงสุดหรือในบางครั้งอาจเลือกโครโมโซมที่มีค่าความเหมาะสมปานกลางและต่ำบางส่วนเข้ามาด้วย เพราะบางกรณีการนำสายพันธุ์ที่มีค่าปานกลางหรือต่ำมาผสมกันจะสามารถทำให้เกิดสายพันธุ์ที่ดีในรุ่นต่อไปได้




**รูปที่ 3.10**แสดงการเลือกโครโมโซมตามค่าความเหมาะสม

การคัดเลือกข้อมูลมีลักษณะเป็นไปตามหลักการที่ว่า การอยู่รอดของสิ่งที่เหมาะสมที่สุด (Survival of the Fittest) [18][19] ถ้าเป็นการวัดค่าความเหมาะสมจากค่าสูงสุด (Maximized Value) ความน่าจะเป็นของของแต่ละโครโมโซม ที่จะได้รับการสุ่มเลือกแต่ละครั้ง (Probability of Value Selection: $P_{Si}$) จะเป็นไปดังสมการที่ (3.11)

$$P_{Si} = \frac{f_i}{\sum_{i=1}^{n} f_i} \tag{3.11}$$

เมื่อ ค่าความเหมาะสมของแต่ละทางเลือก ($f_i$) เทียบกับผลรวมค่าความเหมาะสมทั้งหมด หากเป็นการวัดค่าความเหมาะสมจากค่าต่ำสุด (Minimized Value) ความน่าจะเป็นของแต่ละโครโมโซมที่จะได้รับการสุ่มเลือกแต่ละครั้ง จะเป็นไปดังสมการที่ (3.12)

$$P_{Si} = 1 - \frac{f_i}{\sum_{i=1}^{n} f_i} \tag{3.12}$$

ดังนั้นสามารถคำนวณค่าความคาดหวังที่จะสุ่มได้ (Expected Value : E) ของแต่ละโครโมโซม เป็นไปดังสมการที่ (3.13)

$$E = P_{Si} \times Popsize \tag{3.13}$$

เมื่อ Popsize คือขนาดของประชากรทั้งหมด

ในการสุ่มโครโมโซมของเจเนติกอัลกอริทึมแบบง่าย จะใช้แบบจำลองการหมุนวงล้อถ่วงน้ำหนัก (Roulette Wheel) [19] ซึ่งจะกำหนดขนาดของช่องวงล้อตามความน่าจะเป็นที่จะสุ่มได้ในแต่ละครั้ง ของแต่ละโครโมโซมมีวิธีการดังนี้

1. หาค่าความเหมาะสมของแต่ละโครโมโซม
2. หาความน่าจะเป็นที่จะสุ่มได้ในแต่ละครั้งของแต่ละโครโมโซม
3. หาความถี่สะสม ($q_i$) ของความน่าจะเป็นของแต่ละโครโมโซม ดังสมการที่ (3.14)

$$q_i = \sum_{i=1}^{j} P_{Si} \tag{3.14}$$

4. สร้างเลขสุ่มจำนวนจริง ($r$) ที่มีค่าอยู่ในช่วง [0.0, 1.0]
5. เลือกโครโมโซมลำดับที่ $r$ ซึ่ง $r$ มีค่าอยู่ระหว่าง $q_{i-1}$ และ $q_i$

จากวิธีดังกล่าวจะเห็นได้ว่า โครโมโซมใดที่มีความน่าจะเป็นที่จะถูกเลือกน้อยๆ จะมีโอกาสถูกเลือกขึ้นมาน้อยเพราะช่องว่างระหว่าง $q_{i-1}$ และ $q_i$ จะแคบมาก ทำให้โอกาสที่ r จะตกช่องนั้นมีน้อย ในทางตรงกันข้าม โครโมโซมใดที่มีความน่าจะเป็นสูง ก็จะมีโอกาสถูกเลือกมากเนื่องจากช่องว่างระหว่าง $q_{i-1}$ และ $q_i$ จะกว้าง ซึ่งสอดคล้องกับที่ได้กล่าวไปแล้วว่า ถ้าความน่าจะเป็นที่จะถูกเลือกมีค่ามาก ก็จะมีโอกาสที่จะถูกเลือกไปเป็นประชากรรุ่นใหม่สูงตามไปด้วย ดังแสดงในตารางที่ 2

**ตารางที่ 3.2** ตัวอย่างการใช้แบบจำลองวงล้อถ่วงน้ำหนัก

| โครโมโซม | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|
| ค่าความเหมาะสม | 8 | 2 | 17 | 7 | 2 |
| ค่าความน่าจะเป็นที่จะสุ่มได้แต่ละครั้ง ($P_{Si}$) | 0.22 | 0.06 | 0.47 | 0.19 | 0.06 |
| ความถี่สะสมค่าความน่าจะเป็น ($q_i$) | 0.22 | 0.28 | 0.75 | 0.94 | 1.00 |
| สร้างเลขสุ่มจากการหมุนวงล้อแต่ละครั้ง (r) | 0.33 | 0.84 | 0.45 | 0.12 | 0.28 |
| โครโมโซมที่ถูกเลือก | 3 | 4 | 3 | 2 | 1 |

นอกเหนือจากการคัดเลือกโดยใช้แบบจำลองการหมุนวงล้อถ่วงน้ำหนัก [20] ได้รวบรวมวิธีการอื่นๆ ที่ใช้ในกระบวนการคัดเลือกได้แก่ การคัดเลือกแบบจัดลำดับ (Ranking Selection) การคัดเลือกตามสถานะที่แน่นนอน (Steady-State Selection) และการคัดเลือกแบบมีอภิสิทธิ์ (Elitism Selection) ซึ่งเราสามารถเลือกใช้ได้ตามความเหมาะสมในแต่ละงาน

**3.6.2.6 การครอสโอเวอร์ (Crossover)**

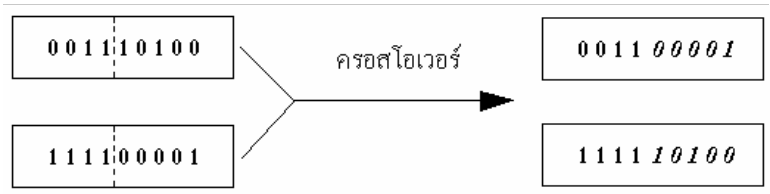
การครอสโอเวอร์ คือการนำโครโมโซม 2 โครโมโซม มาทำการตามขั้นตอนต่างๆ ซึ่งจะให้ค่าโครโมโซมใหม่ที่จะนำไปใช้ในการคัดเลือกครั้งต่อไป หรือหมายถึงการนำโครโมโซมสองโครโมโซมมาผสมกันเพื่อให้ได้ค่าโครโมโซมขึ้นมาใหม่นั่นเอง ในขั้นตอนนี้จะพยายามสร้างทางเลือกใหม่เพื่อเป็นคำตอบให้กับปัญหา และปรับปรุงทางเลือกให้ดีขึ้นโดยการครอสโอเวอร์ ซึ่งเจเนติกอัลกอริทึมจะพยายามสร้างทางเลือกที่ดีขึ้นโดยการรวมลักษณะที่ดีของแต่ละโครโมโซมเข้าด้วยกัน โครโมโซมที่มีค่าความเหมาะสมสูงกว่ามักจะถูกเลือกมาทำการครอสโอเวอร์บ่อยครั้งกว่า ส่งผลให้มีโอกาสในการอยู่รอดไปยังรุ่นต่อไปสูงกว่า การครอสโอเวอร์สามารถทำได้หลายวิธี เช่น การครอสโอเวอร์หนึ่งจุด (One Point Crossover) การครอสโอเวอร์สองจุด (Two Point Crossover) และการครอสโอเวอร์หลายจุด (Multiple Point Crossover) ซึ่งมีวิธีการโดยทั่วไปดังนี้

1. ประชากรทั้งหมดจะถูกนำมาจับคู่โดยการสุ่ม ซึ่งจะได้ผลการจับคู่ออกมาทั้งหมด N/2 คู่ เมื่อ N คือจำนวนประชากรทั้งหมดในรุ่นนั้นๆ
2. สร้างเลขสุ่มจำนวนจริง ($r$) ซึ่งมีค่าอยู่ในช่วง [0.0, 1.0] โดยถ้าจำนวนจริงที่สุ่มได้มีค่าน้อยกว่าค่าความน่าจะเป็นในการครอสโอเวอร์ (Probability of Crossover : $P_c$) แล้วโครโมโซมคู่นั้นจึงจะเกิดการครอสโอเวอร์
3. ครอสโอเวอร์โดยแลกเปลี่ยนส่วนของคู่โครโมโซมพ่อแม่นั้น โดย
   - สุ่มเลือกตำแหน่งที่จะทำการครอสโอเวอร์
   - แลกเปลี่ยนค่าในแต่ละบิทของคู่โครโมโซมพ่อแม่ ตั้งแต่ตำแหน่งที่สุ่มได้จนหมด ซึ่งทำให้เกิดโครโมโซมรุ่นลูกใหม่จำนวน 2 โครโมโซม

การครอสโอเวอร์ในแต่ละรุ่นสามารถทำได้มากกว่า 1 คู่ ขึ้นอยู่กับอัตราค่าความน่าจะเป็นในการครอสโอเวอร์ ซึ่งจำนวนของการครอสโอเวอร์สามารถคำนวณได้ตามสมการที่ 3.15

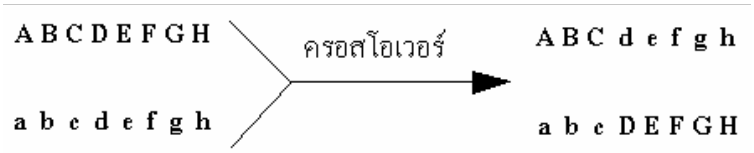
$$\text{จำนวนของการครอสโอเวอร์} = P_c \times (\text{Popsize} / 2) \qquad (3.15)$$

เมื่อ $P_c$ คือความน่าจะเป็นในการครอสโอเวอร์ และ Popsize คือขนาดของประชากรในรุ่นนั้นๆ ตัวอย่างการครอสโอเวอร์แบบไบนารี เช่น ถ้าสุ่มตำแหน่งที่จะทำการครอสโอเวอร์ได้เป็นตำแหน่งที่ 4 การแลกเปลี่ยนส่วนของโครโมโซมจะเกิดขึ้นหลังตำแหน่งที่ 5 เรื่อยไปจนถึงตำแหน่งสุดท้าย เกิดโครโมโซมใหม่ขึ้นมา 2 โครโมโซม ดังแสดงในรูปที่ 3.11



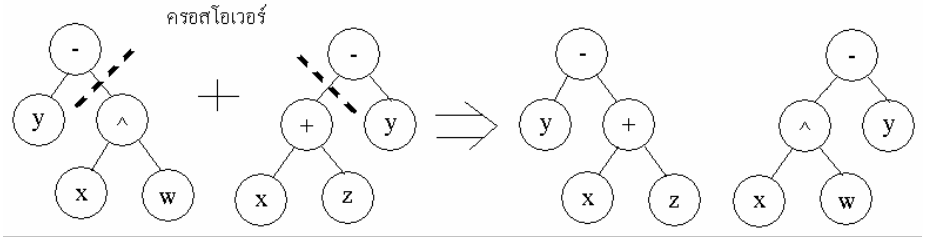

**รูปที่ 3.11** แสดงกระบวนการครอสโอเวอร์แบบไบนารี

ตัวอย่างการครอสโอเวอร์โครโมโซมแบบตัวอักษรหลังตำแหน่งที่ 3 แสดงดังรูปที่ 3.12




**รูปที่ 3.12** แสดงกระบวนการครอสโอเวอร์แบบตัวอักษร

ตัวอย่างการครอสโอเวอร์แบบทรี แสดงดังรูปที่ 3.13




**รูปที่ 3.13** การครอสโอเวอร์ของโครโมโซมแบบทรี

**3.6.2.7 การมิวเตชัน (Mutation)**

การมิวเตชันหรือการผ่าเหล่า เป็นลักษณะของการนำโครโมโซมเก่ามาสุ่มแก้ไขค่าบางค่า เช่น ทำให้ค่าของบางตำแหน่งเปลี่ยนไป โดยทำการกลับบิทเป็นค่าใหม่ในตำแหน่งบิทที่สุ่มได้ ตามค่าความน่าจะเป็นของการมิวเตชันในแต่ละบิท (Probability of Mutation : $P_m$) ที่กำหนด โดยการมิวเตชันจะทำการสุ่มค่า r ของแต่ละตำแหน่งบิทในแต่ละโครโมโซม โดยถ้าค่า r ณ ตำแหน่งของบิทใดเป็นไปดังสมการที่ (3.16)

$$r \leq P_m \tag{3.16}$$

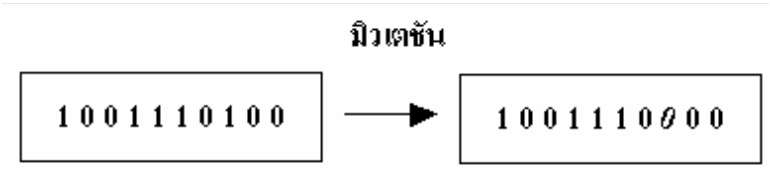
ค่าของบิท ณ ตำแหน่งนั้นก็จะถูกทำมิวเตชัน จำนวนบิทที่จะถูกทำการมิวเตชันนั้นสามารถคำนวณได้ดังสมการที่ (3.17)

จำนวนของการมิวเตชัน = $P_m$ x *Popsize* x $L$ (3.17)

เมื่อ $P_m$ คือ ความน่าจะเป็นของการมิวเตชัน
*Popsize* คือ ขนาดของประชากรในรุ่นนั้นๆ
$L$ คือ ความยาวของโครโมโซม

ผลจากการมิวเตชัน ทำให้ได้โครโมโซมใหม่ที่มีรูปแบบของโครโมโซมแตกต่างจากเดิม ซึ่งมีโอกาสที่จะเป็นโครโมโซมที่ดีขึ้นหรือเลวลงก็ได้ หากโครโมโซมที่ได้ใหม่เป็นโครโมโซมที่เลวลง คือ มีค่าความเหมาะสมต่ำลง โครโมโซมที่ได้นี้ก็จะถูกคัดออกไปในขั้นตอนการคัดเลือก (Selection) นั่นเอง

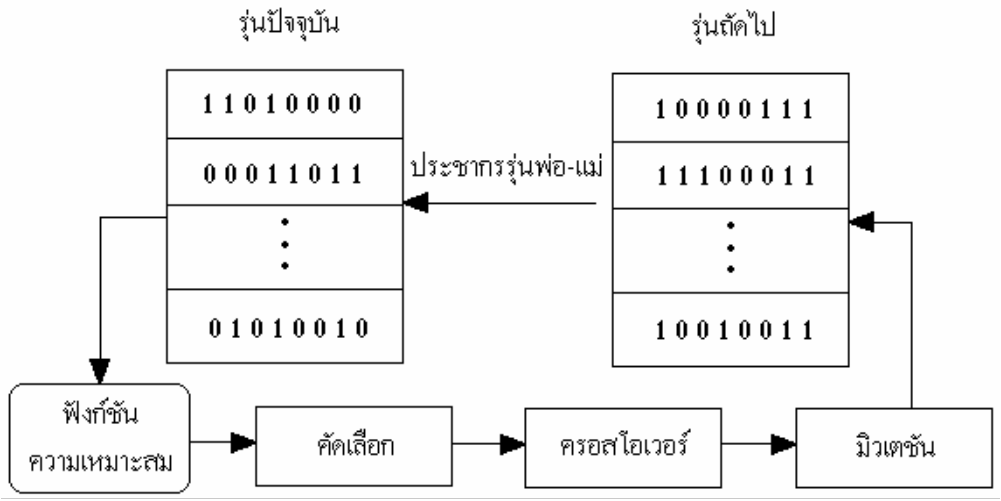
วัตถุประสงค์ของการมิวเตชันคือ เพื่อประกันการสูญหายของข้อมูล และเพื่อความหลากหลายของข้อมูล สำหรับไบนารีมิวเตชัน เป็นการปรับเปลี่ยนข้อมูล ณ ตำแหน่งที่กำหนดนั้น โดยเปลี่ยนข้อมูลจาก 0 เป็น 1 หรือกลับกัน ตัวอย่างของการทำมิวเตชันแสดงดังรูปที่ 3.14 งเป็นการสุ่มเลือกทำมิวเตชัน ณ ตำแหน่งที่ 8




**รูปที่ 3.14** แสดงกระบวนการไบนารีมิวเตชัน

**3.6.2.8 การสร้างประชากรรุ่นใหม่**

ประชากรรุ่นใหม่เป็นกลุ่มโครโมโซมรุ่นลูกที่เกิดจากกระบวนการของวิวัฒนาการต่างๆ ทั้งหมด เริ่มตั้งแต่การวัดค่าความเหมาะสม ทำการคัดเลือก สุ่มเลือกเพื่อนำมาทำครอสโอเวอร์และมิวเตชันตามค่าความน่าจะเป็นที่ได้กำหนดไว้ ซึ่งเมื่อชุดโครโมโซมรุ่นลูกผ่านวิวัฒนาการต่างๆ ดังที่ได้กล่าวไปแล้วนั้นก็จะทำให้เกิดประชากรรุ่นใหม่ โดยที่ประชากรรุ่นใหม่นี้จะถูกถ่ายทอดกลายเป็นประชากรรุ่นเก่า สำหรับวิวัฒนาการรุ่นถัดไปเช่นเดียวกัน ซึ่งจะเรียกวิวัฒนาการนี้ว่า การถ่ายทอดแบบทั่วไป หรือรีโพรดักชันแบบทั่วไป (General Reproduction) ขบวนการทั้งหมดในการสร้างประชากรสามารถแสดงได้ดังรูปที่ 3.15




**รูปที่ 3.15** แสดงกระบวนการสร้างประชากรในรุ่นถัดไป

**3.6.2.9 การกำหนดค่าตัวแปรต่างๆ**

การกำหนดค่าตัวแปรต่างๆ ในกระบวนการเจเนติกอัลกอริทึมนั้น สามารถแบ่งออกได้เป็น 2 กลุ่ม ได้แก่ ตัวแปรที่ใช้ควบคุมการทำงานและเงื่อนไขการสิ้นสุดการทำงาน

**a) ตัวแปรที่ใช้ควบคุมการทำงาน** ตัวแปรในกลุ่มนี้ได้แก่

1. การกำหนดขนาดของประชากร (Population Size) จำนวนขนาดของประชากรมีผลกระทบต่อประสิทธิภาพ ความเร็วในการค้นหาคำตอบ และการใช้ทรัพยากรของระบบ ถ้าจำนวนประชากรน้อยเกินไป อาจทำให้ได้คำตอบที่ขาดประสิทธิภาพในการแก้ปัญหานั้นๆได้ แต่หากจำนวนประชากรมีมากเกินไปแล้วจะส่งผลให้การทำงานเพื่อค้นหาคำตอบจะต้องใช้เวลาและทรัพยากรมากขึ้น ดังนั้นการกำหนดจำนวนขนาดของประชากรจะต้องมีความเหมาะสม [21]

2. การกำหนดค่าความน่าจะเป็นในการคัดเลือก (Probability of Selection) การสุ่มเลขเพื่อเข้าสู่การคัดเลือกโดยใช้แบบจำลองการหมุนวงล้อถ่วงน้ำหนัก (Roulette Wheel) ถ้าตัวเลขที่สุ่มได้ทำให้เกิดช่วงค่าที่แคบเกินไป หรือกว้างเกินไป อาจทำให้โครโมโซมที่ดีไม่ถูกคัดเลือก หรือทำให้เกิดการคัดเลือกโครโมโซมบางตัวซ้ำๆ แม้ว่าในรุ่นนั้นจะมีประชากรโครโมโซมอื่นๆ อีกก็ตาม

3. การกำหนดค่าความน่าจะเป็นในการครอสโอเวอร์ (Crossover Probability) ที่เหมาะสมเพื่อผลิตโครโมโซมที่มีความหลากหลายในประชากรรุ่นต่อไป การกำหนดค่าความน่าจะเป็นในการครอสโอเวอร์ที่เหมาะสมมีส่วนในการเพิ่มประสิทธิภาพในการค้นหาคำตอบ

4. การกำหนดค่าความน่าจะเป็นในการมิวเตชัน (Mutation Probability) ที่เหมาะสม ซึ่งแต่ละปัญหาก็จะต้องการค่าความน่าจะเป็นในการครอสโอเวอร์และมิวเตชันที่แตกต่างกันไป ดังนั้นจึงควรเลือกใช้ให้เหมาะสมกับแต่ละปัญหา เพื่อให้การค้นหาคำตอบมีประสิทธิภาพมากที่สุด

5. จำนวนรุ่น (Number of Generation) ในการค้นหาคำตอบ

**b) เงื่อนไขการสิ้นสุดการทำงาน**

โดยปกติการแก้ไขปัญหาจะเสร็จสิ้นเมื่อได้คำตอบที่ดีที่สุด คือได้โครโมโซมที่มีค่าความเหมาะสมสูงสุด สามารถแก้ไขปัญหาหรือเป็นคำตอบที่ดีที่สุดของปัญหานั้นๆ หรือในแนวทางหนึ่ง สามารถกำหนดให้กระบวนการเจเนติกอัลกอริทึมสิ้นสุดการทำงานเมื่อถึงรุ่นสูงสุด (Max Generation) ที่กำหนดไว้ แล้วนำโครโมโซมที่มีค่าความเหมาะสมสูงสุดมาเป็นคำตอบที่ใกล้เคียง

กล่าวโดยสรุป เจเนติกอัลกอริทึมเป็นเทคนิคที่ใช้ในการค้นหาคำตอบ ซึ่งเลียนแบบมาจากกระบวนการวิวัฒนาการทางธรรมชาติที่นำมาประยุกต์ใช้กับคอมพิวเตอร์ เพื่อช่วยแก้ปัญหาในการหาคำตอบต่างๆ ซึ่งพื้นฐานการทำงานเบื้องต้นเป็นเจเนติกอัลกอริทึมแบบง่าย มีรูปแบบโครโมโซมเป็นแบบไบนารี การคัดเลือกใช้แบบจำลองการหมุนวงล้อถ่วงน้ำหนัก การครอสโอเวอร์เป็นการครอสโอเวอร์แบบหนึ่งจุด และมิวเตชันแบบไบนารี ซึ่งสามารถช่วยแก้ปัญหาในการค้นหาคำตอบให้แก่ระบบได้ ในการประยุกต์ใช้เจเนติกอัลกอริทึมกับปัญหาต่างๆ นั้นจะต้องมีการปรับปรุง เปลี่ยนแปลงในบางส่วน เช่น รูปแบบของโครโมโซม ฟังก์ชันความเหมาะสม หรือค่าตัวแปรต่างๆ เพื่อให้เข้ากับรูปแบบของปัญหา และเพื่อให้สามารถค้นหาคำตอบที่ดีที่สุดให้แก่ปัญหานั้น