บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

ในทศวรรษที่ ค.ศ. 1970 บริษัท IBM ได้นำสื่อบันทึกข้อมูลแบบใหม่ ซึ่งบริษัท IBM เรียกว่า จานแม่เหล็ก (Magnetic Disk) ออกสู่ตลาดเป็นครั้งแรก ต้นกำเนิดของจานแม่เหล็กมีแนวความคิดมาจากดรัมแม่เหล็ก (Magnetic Drum) จานแม่เหล็กมีลักษณะกลมแบน มีเส้นผ่าศูนย์กลางประมาณ 14 นิ้ว ผิวทั้งสองหน้าฉาบด้วยสารแม่เหล็ก บนผิวทั้งสองหน้าจะถูกแบ่งออกแนววงกลมเป็นวง ๆ ที่ไม่ต่อเนื่องกัน ตั้งแต่ริมด้านนอกของแผ่นจนถึงริมด้านในคล้าย ๆ กับจานแผ่นเสียง เรียแนววงกลมแต่ละวงว่า แทรค (Track) ทุก ๆ แทรคที่ตรงกันในแนวดิ่งรวมเรียกว่า ไซลินเดอร์ (Cylinder) และในแต่ละแทรคอาจแบ่งย่อยออกเป็น เซกเมนต์ (Segment) หรือ เซกเตอร์ (Sector) เนื่องจากแทรคต่าง ๆ มีขนาดไม่เท่ากัน แต่ทุก ๆ แทรคจะเก็บข้อมูลได้เป็นจำนวนที่เท่ากัน ดังนั้น ความหนาแน่นของข้อมูลในแทรคที่อยู่ด้านในจะมีค่ามากกว่าความหนาแน่นของข้อมูลในแทรคที่อยู่ด้านนอก การอ่านหรือการบันทึกข้อมูลลงบนจานแม่เหล็ก จะต้องอาศัยหัวอ่าน/บันทึก (Read/Write Head) ขณะที่อ่านหรือบันทึกข้อมูลจานแม่เหล็กจะหมุนด้วยความเร็วคงที่ โดยมีหัวอ่าน/บันทึก อยู่บนผิวด้านบนและด้านล่าง ของผิวหน้าของจานแม่เหล็ก

หลังจากที่บริษัท IBM ได้นำจานแม่เหล็กออกสู่ตลาดเป็นครั้งแรกนั้น ปรากฏว่า เป็นที่สนใจของผู้ใช้เครื่องคอมพิวเตอร์ในขณะนั้นเป็นอย่างมาก จึงทำให้จานแม่เหล็กถูกพัฒนาอย่างต่อเนื่อง ทั้งทางด้าน ฮาร์ดแวร์ (Hardware) และซอฟท์แวร์ (Software) เพื่อให้ผู้ใช้สามารถใช้จานแม่เหล็กได้อย่างมีประสิทธิภาพ การพัฒนาจานแม่เหล็กทางด้านฮาร์ดแวร์ ได้แก่ การทำให้จานแม่เหล็กสามารถบันทึกข้อมูลได้มากขึ้น โดยการฉาบผิวหน้าของจานแม่เหล็กด้วยสารที่เรียกว่า Ultra Thin Metallic แทนการใช้ออกไซด์ของเหล็กเคลือบตามธรรมดา

ทำให้สามารถจัดเรียงจำนวนแทรคได้มากขึ้น ซึ่งเป็นผลให้จานแม่เหล็กสามารถบันทึกข้อมูลได้มากขึ้นกว่าเดิม ส่วนการพัฒนาจานแม่เหล็กทางด้านซอฟท์แวร์นั้น เนื่องจากจานแม่เหล็กเป็นสื่อการบันทึกข้อมูลที่เรียกว่า DASD (Direct Access Storage Device) จึงก่อให้เกิดการพัฒนาทางด้านซอฟท์แวร์เกี่ยวกับการจัดแฟ้มข้อมูลบนจานแม่เหล็ก ทำให้สามารถจัดแฟ้มข้อมูลได้หลายลักษณะ เช่น

- Sequential File
- Indexed Sequential File
- Direct File
- Key Sequenced File
- Entry Sequenced File
- Relative Record File
- Text File
- Compressed File

จากตัวอย่างรูปแบบการจัดแฟ้มข้อมูลที่ได้กล่าวถึงข้างต้น ส่วนใหญ่ขนาดของโลจิกคัล เรคคอร์ด (Logical Record) จะมีความยาวคงที่ (Fixed Length) เท่ากันทุก ๆ เรคคอร์ด ข้อดีคือตำแหน่งของโลจิกคัล เรคคอร์ดจะอยู่คงที่แน่นอน แต่มีข้อเสีย คือ ในกรณีที่แฟ้มข้อมูลมีขนาดใหญ่จะต้องใช้เนื้อที่ของจานแม่เหล็กเป็นจำนวนมาก ถ้าหากโลจิกคัล เรคคอร์ด มีขนาดความยาวไม่คงที่ (Variable Length) จะทำให้การใช้เนื้อที่ในจานแม่เหล็กสำหรับบันทึกข้อมูลน้อยลง ซึ่งมีเทคนิคการทำได้หลายวิธี แตกต่างกันไป การจัดแฟ้มข้อมูลแบบอัด (Compressed File) เป็นเทคนิคการจัดแฟ้มข้อมูลแบบหนึ่งที่ทำชื่อเสียงให้กับบริษัทคาดาพอยต์เป็นอย่างมาก ผู้วิจัยได้ทำการศึกษาถึงวิธีการอัดข้อมูลของเครื่องคาดาพอยต์ พบว่า การอัดข้อมูลที่เป็นที่ว่าง (Space) นั้น จะประกอบด้วยไบท์ที่ใช้ในการควบคุม (Control Byte) จำนวน 1 ไบท์ มีค่าเป็น 011 ฐานแปด เพื่อใช้บอกจุดเริ่มต้นของการอัดข้อมูลที่เป็นที่ว่าง โดยอีก 1 ไบท์ ที่ตามมาจะใช้เป็นตัวนับแบบเลขฐานสอง (Binary Count Byte) เพื่อบอกจำนวนที่ว่างที่ถูกอัด เมื่อมีที่ว่างอยู่ติดต่อกันเป็นจำนวนตั้งแต่ 2 ตัวขึ้นไป จนถึง 255 ตัว จะถูกแทนด้วยจำนวนไบท์เพียง 2 ไบท์ ดังที่

กล่าวมาแล้ว ทำให้การใช้เนื้อที่บันทึกข้อมูลในจานแม่เหล็กลดลง เครื่องคาดคาพยค์ 5500 ที่ติดตั้งอยู่ ณ สถาบันบริการคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย ได้ถูกใช้งานทางด้านการเตรียมข้อมูลเสียเป็นส่วนใหญ่ และข้อมูลเกือบทั้งหมดที่เตรียมจะเป็นตัวเลข ดังนั้น กรณีที่ตัวเลขซ้ำกันหลาย ๆ ตัว เช่น เลข 0 ในช่องจำนวนเงิน จะไม่ถูกทำการอัด เพื่อปรับปรุงวิธีการอัดข้อมูลของเครื่องคาดคาพยค์ โดยทำให้สามารถอัดข้อมูลที่ซ้ำกันหลาย ๆ ตัวได้ โดยผู้วิจัยพิจารณาเห็นว่า ควรจะแยกการอัดข้อมูลเป็น 2 ลักษณะ คือ การอัดข้อมูลที่เป็นที่ว่าง และการอัดข้อมูลที่ไม่ใช่ที่ว่าง การอัดข้อมูลที่เป็นที่ว่าง จะยังคงรูปแบบเดิมของเครื่องคาดคาพยค์ตามที่ได้กล่าวมาแล้วข้างต้น ส่วนการอัดข้อมูลที่ไม่ใช่ที่ว่าง จะใช้จำนวนไบท์ 3 ไบท์ โดยไบท์แรกมีค่าเป็น 013 ฐานแปด เพื่อบอกจุดเริ่มต้นของการอัดข้อมูลที่ไม่ใช่ที่ว่างอีก 1 ไบท์ ที่ตามมาจะเก็บค่าของตัวเลข หรือตัวอักษรที่ถูกทำการอัดและอีก 1 ไบท์ที่เหลือจะใช้เป็นตัวนับแบบเลขฐานสองว่าจำนวนตัวเลขหรือตัวอักษรที่ถูกอัดมีเป็นจำนวนเท่าใด ดังนั้นเมื่อมีจำนวนตัวเลขหรือตัวอักษรซ้ำกันตั้งแต่ 3 ตัวขึ้นไป จนถึง 255 ตัว จะถูกแทนด้วยจำนวนไบท์เพียง 3 ไบท์ เท่านั้น ซึ่งจะทำให้ประหยัดเนื้อที่บันทึกข้อมูลในจานแม่เหล็กได้เป็นอย่างมาก

1.2 วัตถุประสงค์และขอบเขตการวิจัย

ในการทำวิจัยนี้จะทำการศึกษาถึงลักษณะการจัดแฟ้มข้อมูลแบบอัดของเครื่องคาดคาพยค์ รวมทั้งศึกษาโปรแกรมต่างๆ ที่เกี่ยวข้องกับแฟ้มข้อมูลแบบอัด จากนั้น จะทำการสร้างโปรแกรมที่ใช้ในการจัดรูปแบบของแฟ้มข้อมูลแบบอัดและโปรแกรมอื่น ๆ ที่เกี่ยวข้องกับแฟ้มข้อมูลแบบอัด คือโปรแกรมที่ใช้ในการจัดรูปแบบของแฟ้มข้อมูลแบบอัด โปรแกรมที่ใช้แสดงเนื้อหาของแฟ้มข้อมูลแบบอัด และโปรแกรมที่ใช้ถ่ายเทข้อมูลระหว่างจานแม่เหล็ก และเทปแม่เหล็ก โดยจะทำการทดสอบโปรแกรมที่สร้างขึ้น ว่าจะสามารถทำงานได้จริงตามความต้องการ

1.3 ขั้นตอนและวิธีดำเนินการวิจัย

1.3.1 ศึกษาการจัดแฟ้มข้อมูลแบบอัดของเครื่องคาดคาพยุต์

1.3.2 ศึกษาภาษาแอสเซมบลีของเครื่องคาดคาพยุต์

1.3.3 ศึกษาการทำงานของโปรแกรมที่เกี่ยวข้องกับแฟ้มข้อมูลแบบอัด

1.3.4 ทำการแปลงโปรแกรมที่เกี่ยวข้องกับการจัดแฟ้มข้อมูลแบบอัดจากภาษาเครื่องมาเป็นภาษาแอสเซมบลีของเครื่องคาดคาพยุต์

1.3.5 สร้างโปรแกรมที่ใช้ในการจัดรูปแบบของแฟ้มข้อมูลแบบอัดแบบใหม่ โปรแกรมที่ใช้ในการสร้างแฟ้มข้อมูลแบบอัด โปรแกรมที่ใช้ในการแสดงเนื้อหาของแฟ้มข้อมูลแบบอัด และโปรแกรมที่ใช้สำหรับถ่ายข้อมูลระหว่างจานแม่เหล็กและเทปแม่เหล็ก

1.3.6 ทดสอบโปรแกรมที่สร้างขึ้นใหม่ทั้งหมด ตามข้อ 1.3.5

1.3.7 สรุปผลการวิจัย

1.4 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1.4.1 ทำให้ทราบถึงลักษณะการจัดแฟ้มข้อมูลแบบอัดของเครื่องคาดคาพยุต์

1.4.2 โปรแกรมที่สร้างขึ้นสามารถนำไปใช้กับเครื่องคาดคาพยุต์ได้

1.4.3 แฟ้มข้อมูลแบบอัดแบบเดิมสามารถนำไปใช้ได้กับโปรแกรมที่สร้างขึ้นใหม่ โดยไม่ต้องมีการเปลี่ยนแปลงใดๆ ก่อนนำไปใช้

1.4.4 ทำให้สามารถใช้เนื้อที่ในจานแม่เหล็กได้อย่างมีประสิทธิภาพมากขึ้นกว่าเดิม

1.4.5 เป็นแนวทางสำหรับผู้ที่สนใจจะทำการศึกษาต่อไป