บทที่ ๔

การทำงานของคำสั่ง 'SORT'

โดยปกติแล้วในระบบงานทางธุรกิจ สิ่งที่จำเป็นสำหรับระบบคือ ข้อมูลที่เรียงลำดับตามเขตข้อมูลที่ใช้งาน และในการเขียนโปรแกรมหนึ่ง ๆ อาจมีความจำเป็นในการเรียงลำดับข้อมูลภายในโปรแกรม ซึ่งแต่เดิมเราไม่สามารถทำได้กับอินเตอร์เพรตเตอร์ตัวเก่า จึงได้พัฒนาคำสั่งที่ใช้ในการเรียงลำดับข้อมูลภายในโปรแกรมขึ้นมาเพื่อใช้งาน เรียกคำสั่งนี้ว่า 'SORT'

๔.๑ การจัดเรียงลำดับข้อมูล

การจัดเรียงลำดับข้อมูล คือ การเรียงลำดับข้อมูลจากแฟ้มข้อมูลหนึ่ง ได้เป็นอีกแฟ้มข้อมูลหนึ่ง ที่มีการจัดเรียงลำดับข้อมูลตามเขตข้อมูลที่ต้องการ สามารถเรียงจากค่าน้อยไปค่ามากหรือเรียงจากค่า มากไปค่าน้อยแล้วแต่ความต้องการ และทำได้ครั้งละหลายเขตข้อมูล แต่จะต้องเป็นชนิดเดียวกันสำหรับคำสั่งที่พัฒนาขึ้นมานี้

๔.๒ ตำแหน่งของระเบียน

คือ ตำแหน่งของระเบียนทางกายภาพของแฟ้มข้อมูลนำเข้า จำเป็นต้องเก็บควบคู่กับเขตข้อมูลที่เรียงลำดับในแฟ้มคีย์ เพื่อประโยชน์ในการบันทึกแฟ้มข้อมูลนำออกนั่นเอง ตำแหน่งของระเบียนจะพราบได้ในขณะที่ทำการอ่านข้อมูลจากแฟ้มข้อมูลนำเข้า

๔.๒.๑ แฟ้มข้อมูลในระบบซีพีเอ็ม มีการแบ่งข้อมูลออกเป็นระเบียน แต่ละระเบียนมีความยาวคงที่คือ ๑๒๘ ไบต์ เรียกว่า ระเบียนกายภาพ (Physical Record) การอ่านหรือบันทึกข้อมูล (Acccess) กระทำครั้งละ ๑ ระเบียนกายภาพ มีอยู่ ๒ วิธี คือ

๔.๒.๑.๑ แบบอันดับ การอ่านหรือบันทึกข้อมูลจะกระทำตั้งแต่ต้นแฟ้มข้อมูล เรียงลำดับไปทีละระเบียน

๔.๒.๑.๒ แบบสุ่ม การอ่านหรือบันทึกข้อมูลสามารถทำที่ระเบียนใดระเบียนหนึ่ง โดยระบุหมายเลขระเบียน

เนื่องจากการบันทึกหรืออ่านข้อมูลจากจานแม่เหล็กแบบอ่อน จะต้องกระทำโดยผ่านเอฟซีบี (File Control Block) ซึ่งเก็บอยู่ในส่วนที่เรียกว่าไดเร็คทอรี (Directory) ของแต่ละแผ่นจานแม่เหล็กแบบอ่อน

๑ เอฟซีบี ใช้ที่ ๓๖ ไบต์ มีรายละเอียดดังนี้

ไบต์ที่ ๑ เก็บตู้จานแม่เหล็กแบบอ่อน

ไบต์ที่ ๒ - ๙ เก็บชื่อของแฟ้มข้อมูล

ไบต์ที่ ๑๐ - ๑๒ เก็บประเภทของแฟ้มข้อมูล

ไบต์ที่ ๑๓ เก็บเอกซ์เทนท์ที่กำลังทำงาน

ไบต์ที่ ๑๔ - ๑๕ จองไว้สำหรับระบบ

ไบต์ที่ ๑๖ เก็บอันดับที่ของระเบียนในเอกซ์เทนท์นั้น ๆ

ไบต์ที่ ๑๗ - ๓๒ จองไว้สำหรับระบบ

ไบต์ที่ ๓๓ เก็บค่าระเบียนกายภาพปัจจุบันที่จะอ่านหรือบันทึกแบบอันดับ

ไบต์ที่ ๓๔ - ๓๕ เก็บค่าระเบียนกายภาพปัจจุบันที่จะอ่านหรือบันทึกแบบสุ่ม

ไบต์ที่ ๓๖ เก็บค่าระเบียนกายภาพปัจจุบันที่จะอ่านหรือบันทึก แบบสุ่มที่เกินกว่าค่าที่จะเก็บได้ในไบต์ที่ ๓๔-๓๕

ในการอ่านหรือบันทึกข้อมูล ถ้าจะกระทำตั้งแต่ระเบียนแรก ก็ต้องกำหนดค่าในไบต์ที่ ๑๓ เป็น 0 การอ่านหรือบันทึกจะกระทำตั้งแต่เอกซ์เทนท์ที่ ๑ และ ได้ระเบียนที่ ๑ เป็นต้นมา ค่าอันดับที่ของระเบียนจะถูกเก็บไว้ในไบต์ที่ ๑๖ และค่านี้จะถูกเพิ่มขึ้นทีละ ๑ เมื่อมีการอ่านหรือบันทึกแบบอันดับ ค่านี้จะถูกเพิ่มขึ้นจนกว่าจะเกินค่าที่เก็บได้ใน ๑ เอกซ์-เทนท์ (คือเกินกว่า ๖๔) แล้วเอกซ์เทนท์ที่ ๒ จะถูกเรียกมาใช้งาน และค่าในไบต์ที่ ๑๖ จะถูกกำหนดเป็น ๑ อีกครั้งหนึ่ง (เป็นอันดับที่ ๑ ในเอกซ์เทนท์ที่ ๒) เช่นนี้จนหมดระเบียนในแฟ้มบนจานแม่เหล็กแบบอ่อน แต่จากค่าเอกซ์เทนท์ที่และอันดับที่ของระเบียนในเอกซ์เทนท์นั้น เราสามารถนำมาหาค่าอันดับที่ของระเบียนในแฟ้มข้อมูลได้ โดยค่านี้จะถูก

เก็บไว้ในไบต์ที่ ๓๔-๓๕ ดังนั้นเพื่อให้ทราบถึงอันดับที่ของระเบียนเราจะต้องเก็บค่านี้ไว้

๔.๒.๒ แฟ้มข้อมูลในระบบอินเตอร์เพรตเตอร์ แบ่งออกเป็น ๒ แบบ คือ

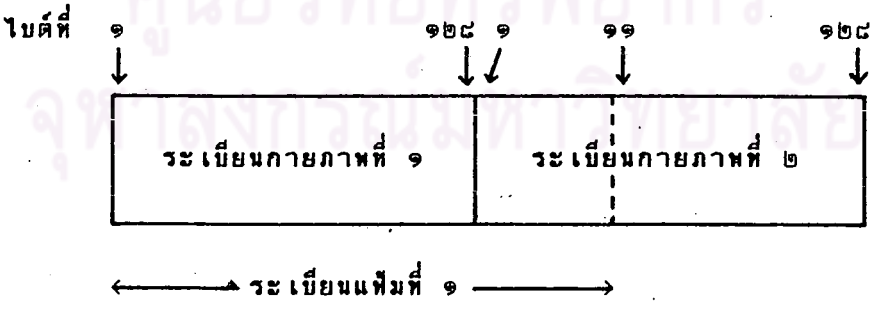
๔.๒.๑.๑ แบบอันดับ แต่ละระเบียนและเขตข้อมูลมีความยาวไม่คงที่ ดังนั้นจะมีเครื่องหมาย '2C' เป็นตัวแยกเขตข้อมูล และเครื่องหมาย 'ODOA' เป็นตัวแยกระเบียน นอกจากนี้จะมีเครื่องหมาย '1A' ปิดท้ายแฟ้มข้อมูลเพื่อบอกการหมดของแฟ้มข้อมูล การอ่านหรือบันทึกข้อมูลจะกระทำทีละระเบียนตั้งแต่ต้นแฟ้มข้อมูล

๔.๒.๑.๒ แบบสุ่ม แต่ละเขตข้อมูลและระเบียนมีความยาวคงที่ ดังนั้นไม่จำเป็นต้องมีเครื่องหมายพิเศษใดๆ เป็นตัวแยกเขตข้อมูลและระเบียน

เรียกระเบียนของแฟ้มข้อมูลในระบบอินเตอร์เพรตเตอร์ว่า ระเบียนแฟ้ม (Logical Record)

๔.๒.๓ ความสัมพันธ์ระหว่างแฟ้มข้อมูลในระบบซีพีเอ็มและแฟ้มข้อมูลในระบบอินเตอร์เพรตเตอร์

เนื่องจากระเบียนกายภาพมีความยาว ๑๒๘ ไบต์ ดังนั้น ถ้าระเบียนแฟ้มที่ ๑ ยาว ๑๓๘ ไบต์ สำหรับระเบียนที่ ๒ ของแฟ้มอาจจะเป็นค่าระเบียนกายภาพที่ ๒ แต่อยู่ในไบต์ที่ ๑๑ ดังรูปที่ ๔.๑



รูปที่ ๔.๑ แสดงความสัมพันธ์ระหว่างระเบียนกายภาพและระเบียนแฟ้ม

ดังนั้นค่าที่เราจะต้องเก็บ เพื่อให้ทราบถึงตำแหน่งของระเบียนแฟ้มในระเบียนกายภาพ นอกจากอันดับที่ของระเบียนแบบสุ่ม คืออันดับที่ของไบต์ในระเบียนกายภาพ

สรุปเพื่อให้ทราบถึงตำแหน่งของระเบียนแฟ้มในระเบียนกายภาพ จะต้องเก็บค่าระเบียนกายภาพ ๒ไบต์ (จากไบต์ที่ ๓๔-๓๕ ในเอฟซีบี) และอันดับที่ของไบต์ในระเบียนกายภาพ ๑ ไบต์ (จากการค้นหาในโปรแกรม) รวม ๓ ไบต์

๔.๓ หลักการทำงานของคำสั่ง 'SORT'

หลักการทำงานของคำสั่ง 'SORT' มีดังต่อไปนี้

๔.๓.๑ ตรวจสอบกฎเกณฑ์ของการเขียนคำสั่ง ถ้าถูกต้องครบถ้วนก็ทำงานขั้นต่อไป

๔.๓.๒ ล้างขยะ โดยการปรับตารางตัวแปรและตารางค่าสตริงทำให้เนื้อที่ว่างที่ใช้งานได้ในหน่วยความจำหลัก (ดูรูป ๒.๔) ที่จะทำเป็นเนื้อที่ทำงาน (Sort Work Area) มากที่สุดเท่าที่จะมากได้ และตรวจสอบขนาดของหน่วยความจำหลักที่จะเป็นเนื้อที่ทำงานให้สามารถเรียงลำดับได้อย่างน้อยที่สุด ๑๐ ข้อมูล พร้อมทั้งคำนวณจำนวนระเบียนที่มากที่สุดที่จะเรียงลำดับในหน่วยความจำหลักได้ในครั้งหนึ่ง ๆ

๔.๓.๓ นำเขตข้อมูลของระเบียนที่จะเรียงลำดับจากแฟ้มข้อมูลนำเข้า พร้อมทั้งตำแหน่งของระเบียนนั้น ๆ ลงเนื้อที่ทำงานในหน่วยความจำหลักจำนวนมากที่สุดที่จะลงได้ตามที่คำนวณได้จาก ๔.๓.๒

เขตข้อมูลของระเบียนที่จะเรียงลำดับและตำแหน่งของระเบียนจะรวมเรียกเป็น 'คีย์'

๔.๓.๔ แล้วเรียงลำดับคีย์ในหน่วยความจำหลักแบบน้อยไปมาก

๔.๓.๕ ถ้าเป็นครั้งที่ ๑ จะบันทึกคีย์ที่เรียงลำดับแล้วขึ้นจานแม่เหล็กแบบอ่อนที่ใช้ทำงาน (Work Diskette) บนแฟ้มคีย์ที่ ๑ (Key File ๑) แล้วทำ ๔.๓.๗

๔.๓.๖ ถ้าไม่ใช่ครั้งที่ ๑ จะทำการผสม (Merge) คีย์ในหน่วยความจำหลักเข้ากับแฟ้มคีย์ที่ ๑ ขึ้นบนแฟ้มคีย์ที่ ๒ (Key File ๒) ในจานแม่เหล็กแบบอ่อนที่ใช้ทำงาน และลบแฟ้มคีย์ที่ ๑ แล้วเปลี่ยนชื่อแฟ้มคีย์ที่ ๒ เป็นแฟ้มคีย์ที่ ๑

ทำซ้ำ ๔.๓.๓ - ๔.๓.๖ จนหมดระเบียนในแฟ้มข้อมูลนำเข้า จะได้แฟ้มคีย์ของคีย์ที่เรียงลำดับแล้วอยู่ในแฟ้มคีย์ที่ ๑

๔.๓.๗ ดึงข้อมูลของตำแหน่งของระเบียน จากคีย์ในแฟ้มคีย์ที่ ๑ ลงหน่วยความจำหลักจนหมด ตั้งแต่คีย์แรกจนถึงคีย์สุดท้าย แล้วลบแฟ้มคีย์ที่ ๑

๔.๓.๘ สร้างแฟ้มข้อมูลนำออก (Output File) จากแฟ้มข้อมูลนำเข้าด้วยตำแหน่งของระเบียนในหน่วยความจำหลัก ตามประเภทของการจัดเรียงลำดับข้อมูลว่าเรียงลำดับจากน้อยไปมากหรือมากไปน้อย คือ ถ้าเรียงลำดับจากน้อยไปมาก ก็สร้างตามลำดับข้อมูลที่อยู่ในหน่วยความจำหลักจากข้อมูลแรกจนถึงข้อมูลสุดท้าย แต่ถ้าเรียงลำดับจากมากไปน้อย จะสร้างตามลำดับข้อมูลที่อยู่ในหน่วยความจำหลัก จากข้อมูลสุดท้ายมายังข้อมูลแรก

รายละเอียดขั้นตอนการทำงานแสดงไว้ในผังงานที่ ๔.๑ และ โปรแกรมการทำงานแสดงไว้ในภาคผนวก ค.

วิธีการจัดเรียงลำดับข้อมูลใช้วิธี 'Quick Sort' ซึ่งเป็นวิธีที่ค่อนข้างรวดเร็วสำหรับข้อมูลที่มีจำนวนมาก และใช้กันแพร่หลายในปัจจุบัน ส่วนการผสมใช้วิธี 'Simple Merge' ซึ่งเป็นวิธีที่ค่อนข้างสะดวกในการทำงาน

๔.๔ การเรียงลำดับข้อมูลด้วยวิธี 'Quick Sort'

เป็นการเรียงลำดับโดยแบ่งคีย์เป็น ๒ พวก คือพวกที่มีค่าคีย์มากกว่า และพวกที่มีค่าคีย์น้อยกว่าคีย์ตัวหนึ่งที่เป็นคีย์อ้างอิง (Reference Key)

วิธีการคือ ครั้งแรกจะนำคีย์แรกมาเป็นคีย์อ้างอิง แล้วทำการหาค่าคีย์ที่น้อยกว่าจากทางท้ายแฟ้ม นำคีย์ที่น้อยกว่าไปแทนที่ตำแหน่งของคีย์อ้างอิงแต่แรกแล้วทำการหาค่าคีย์ที่มากกว่าจากต้นแฟ้มนับตั้งแต่ถัดจากคีย์ที่มีค่าน้อยกว่าถูกย้ายมา ทำเช่นนี้จนไม่มีการหาจะได้คีย์เป็น ๒ พวก คือพวกที่มีค่าคีย์มากกว่าและพวกที่มีค่าคีย์น้อยกว่าคีย์อ้างอิง แล้วใช้วิธีการเดิมสำหรับแต่ละแฟ้มข้อมูลย่อย จนในที่สุด จะได้ข้อมูลที่เรียงลำดับตามต้องการ

ดังแสดงในผังงานที่ ๔.๒

๔.๕ การผสมข้อมูลด้วยวิธี 'Simple Merge'

เป็นการผสมข้อมูลที่เรียงลำดับแล้ว ๒ ชุด ให้เป็นข้อมูลที่เรียงลำดับ ๑ ชุด

วิธีการคือ เปรียบเทียบค่าคีย์ที่ละตำแหน่ง ในข้อมูลทั้ง ๒ ชุด จากต้นแฟ้ม บันทึกค่าคีย์ที่น้อยกว่าไปยังแฟ้มข้อมูลที่ ๓ ทำเช่นนี้จนหมดข้อมูลทั้ง ๒ ชุด

ดังแสดงในผังงานที่ ๔.๓

๔.๖ รูปแบบและการใช้งานของคำสั่ง 'SORT'

สำหรับการใช้คำสั่ง 'SORT' ในการเรียงลำดับข้อมูล ผู้ใช้ไม่จำเป็นต้องเปิดแฟ้มข้อมูลที่จะใช้ แต่ถ้าผู้ใช้ได้ใช้แฟ้มข้อมูลนั้นเป็นแฟ้มข้อมูลนำออกมาก่อนหน้า ผู้ใช้จะต้องปิดแฟ้มข้อมูลนั้นเสียก่อนใช้คำสั่ง 'SORT' นี้ แฟ้มข้อมูลที่ต้องการมาเรียงลำดับจะเรียกว่าแฟ้มข้อมูลนำเข้า และหลังจากเรียงลำดับข้อมูลแล้ว จะได้แฟ้มข้อมูลนำออกที่เรียงลำดับข้อมูลตามต้องการ

รูปแบบการใช้คำสั่ง 'SORT' มีดังนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

SORT {S / R}, d1 : input file, d2 : output file {A / D}(#n, Ln,...)[RL] /

อธิบาย S คือ ประเภทของแฟ้มข้อมูลนำเข้าเป็นแบบอันดับ

R คือ ประเภทของแฟ้มข้อมูลนำเข้าเป็นแบบสุ่ม

d1 คือ ตู้จานแม่เหล็กแบบอ่อนของแฟ้มข้อมูลนำเข้า

d2 คือ ตู้จานแม่เหล็กแบบอ่อนของแฟ้มข้อมูลนำออก

A คือ วิธีเรียงลำดับข้อมูลแบบน้อยไปมาก

D คือ วิธีเรียงลำดับข้อมูลแบบมากไปน้อย

#n คือ ตำแหน่งของเขตข้อมูลที่จะเรียงลำดับไม่เกิน ๔ เขตข้อมูล เรียงลำดับจากเขตข้อมูลหลักไปยังเขตข้อมูลรอง

Ln คือ ความยาวของเขตข้อมูลนั้น

RL คือ ความยาวของแต่ละระเบียนของข้อมูล ในกรณีแฟ้มข้อมูลนำเข้าแบบสุ่ม ไม่เกิน ๒๕๖ ไบต์

/ คือ การบอกตู้จานแม่เหล็กแบบอ่อนที่ใช้ทำงานเดียวกับตู้จานแม่เหล็กแบบอ่อนของแฟ้มข้อมูลนำเข้า ถ้าไม่มีจะใช้ตู้จานแม่เหล็กแบบอ่อนที่ใช้ทำงานคนละตู้กับตู้จานแม่เหล็กแบบอ่อนของแฟ้มข้อมูลนำเข้า

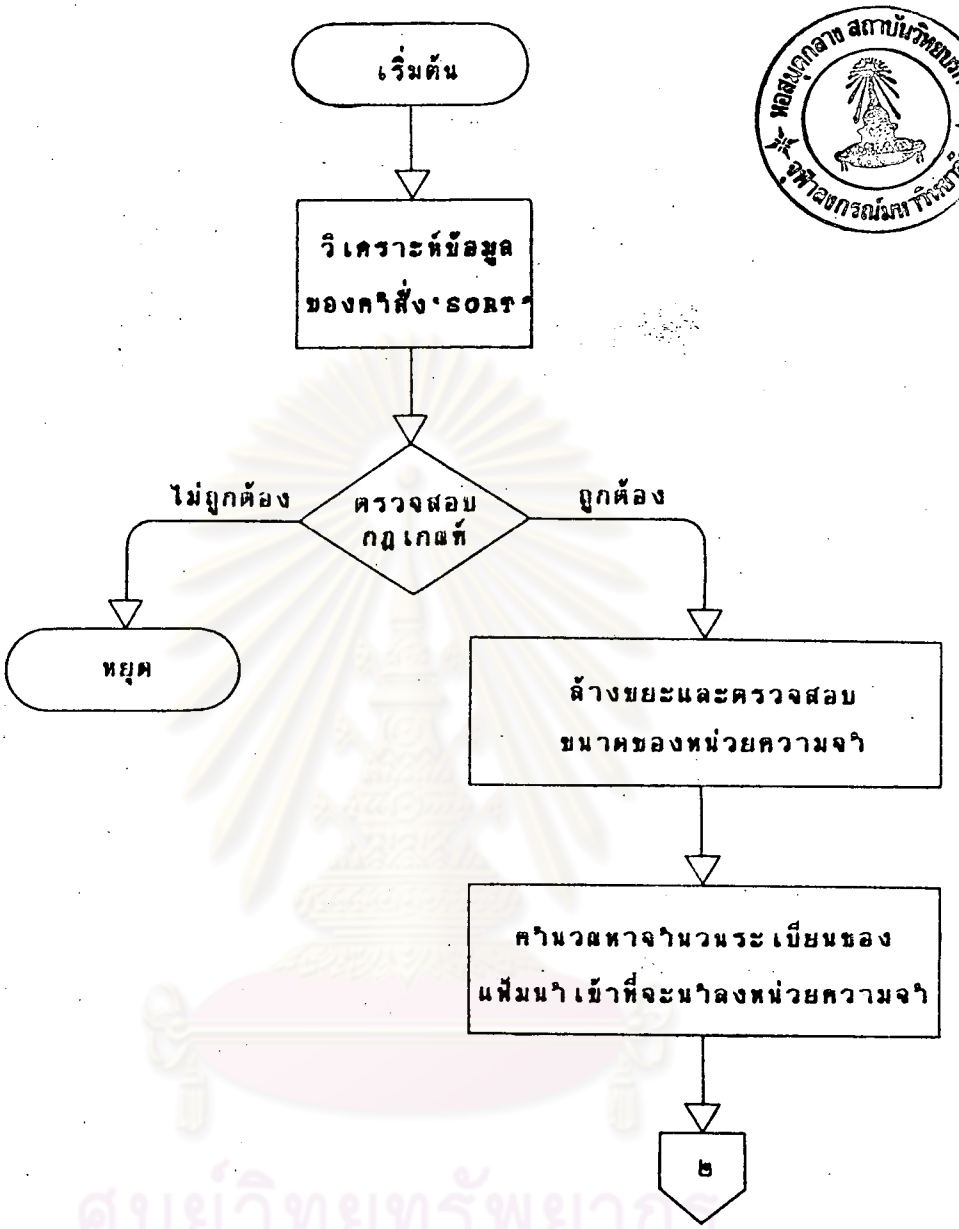
ในการทำงานจำเป็นต้องมีหน่วยความจำสำรองภายนอก ใช้เป็นเนื้อที่ทำงาน คือ บนจานแม่เหล็กแบบอ่อน ตู้จานแม่เหล็กแบบอ่อนที่จะใช้ คือตู้ที่ไม่ใช่ที่กำหนดเป็นที่อยู่ของแฟ้มข้อมูลนำเข้า เช่น ถ้าแฟ้มข้อมูลนำเข้าอยู่ที่ตู้จานแม่เหล็กแบบอ่อน A แล้วเนื้อที่จานแม่เหล็กแบบอ่อนที่ใช้ทำงานจะอยู่ที่ตู้จานแม่เหล็กแบบอ่อน B เป็นต้น

ขนาดของเนื้อที่จานแม่เหล็กแบบอ่อนที่ใช้ทำงาน ที่จะต้องใช้ในการเรียงลำดับแฟ้มข้อมูลหนึ่ง ๆ คำนวณดังนี้

ขนาดของเนื้อที่จานแม่เหล็กแบบอ่อนที่ใช้ทำงาน = จำนวนระเบียนทั้งหมด *

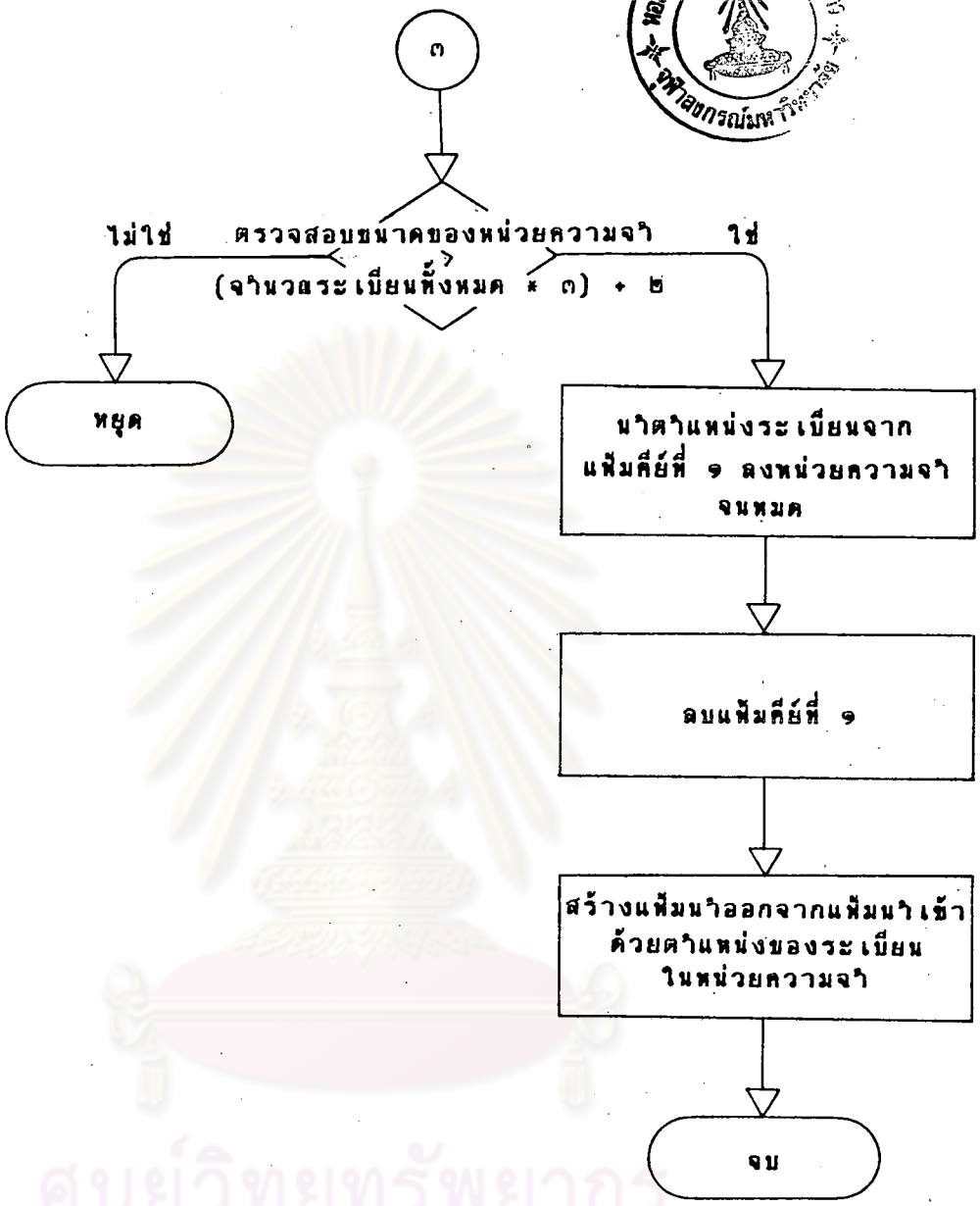
(ความยาวของคีย์รวม + ๓)

เช่น จากตัวอย่างในภาคผนวก ง. ข้อมูลทั้งหมด ๑๕๖ ระเบียน มีความยาวรวม ๓๓ ไบต์ ดังนั้น ขนาดเนื้อที่จานแม่เหล็กแบบอ่อนที่ใช้ทำงาน = ๑๕๖ * (๓๓ +๓) = ๕,๔๖๐ ไบต์ หรือประมาณ ๕.๔ กิโลไบต์ เป็นต้น



ผังงานที่ ๔.๑ แสดงการทำงานของคำสั่ง 'SORT'

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



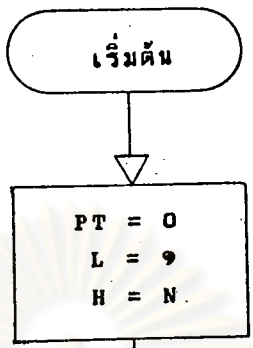
ผังงานที่ ๔.๑ แสดงการทำงานของคำสั่ง 'SORT' (ต่อ)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

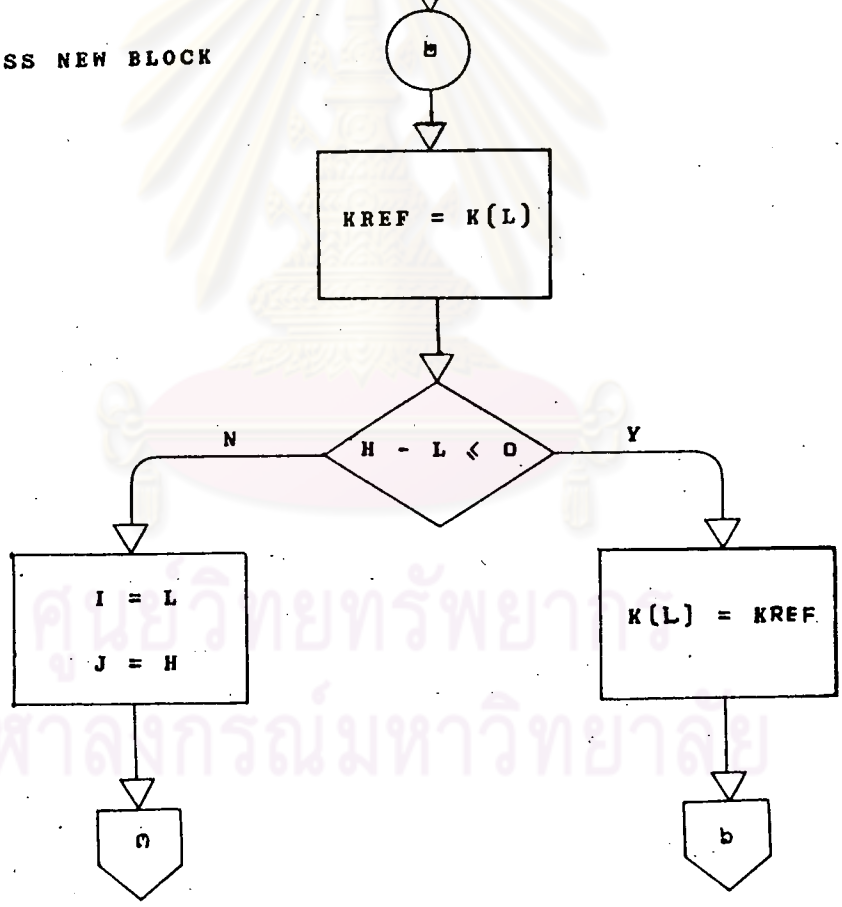
K(i) = คีย์ที่ i ของแฟ้มคีย์

N = จำนวนคีย์ทั้งหมดของแฟ้มคีย์

๑. INITIAL-PROCESS



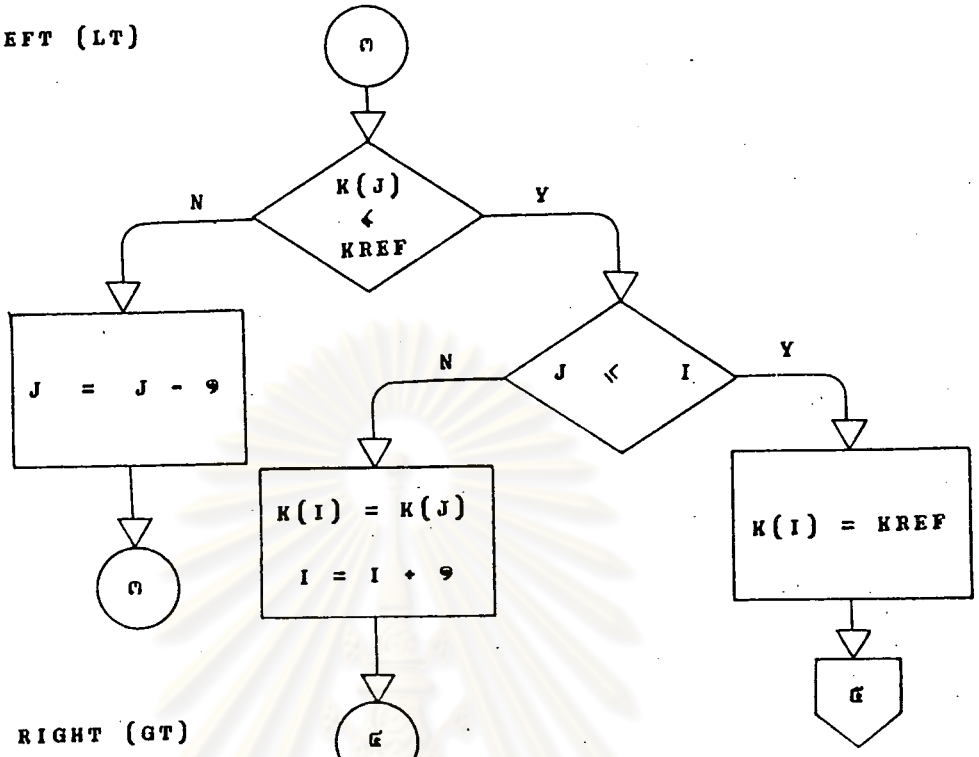
๒. PROCESS NEW BLOCK



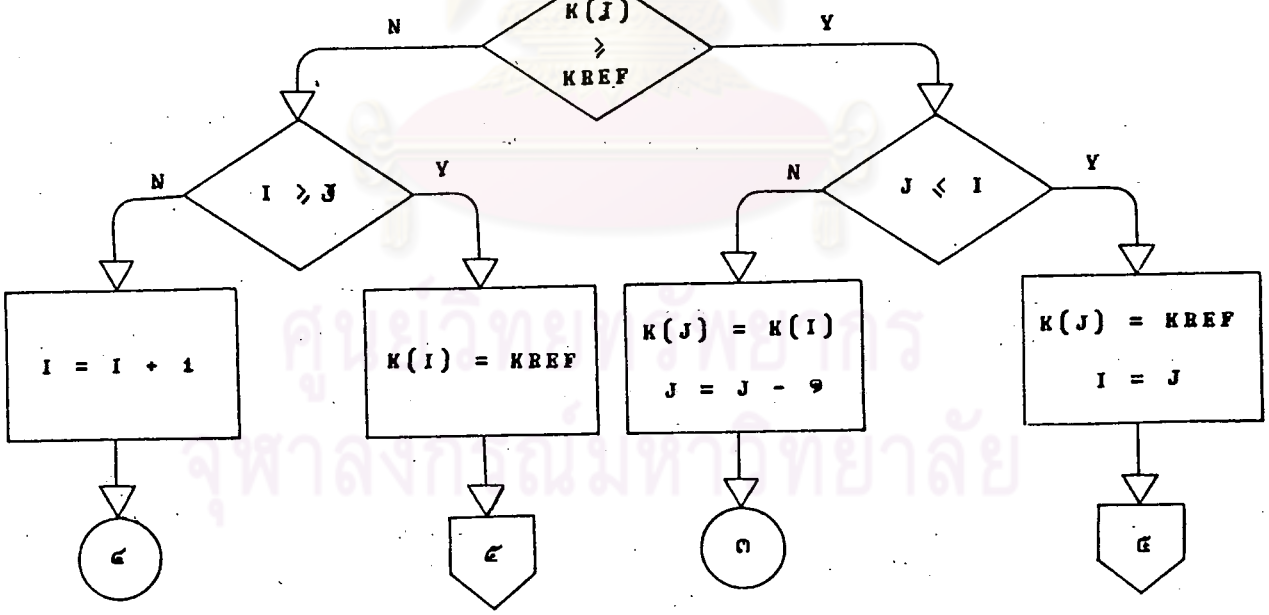
ผังงานที่ ๔.๒ แสดงการเรียงลำดับด้วยวิธี 'QUICK SORT'

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

๓. RIGHT TO LEFT (LT)

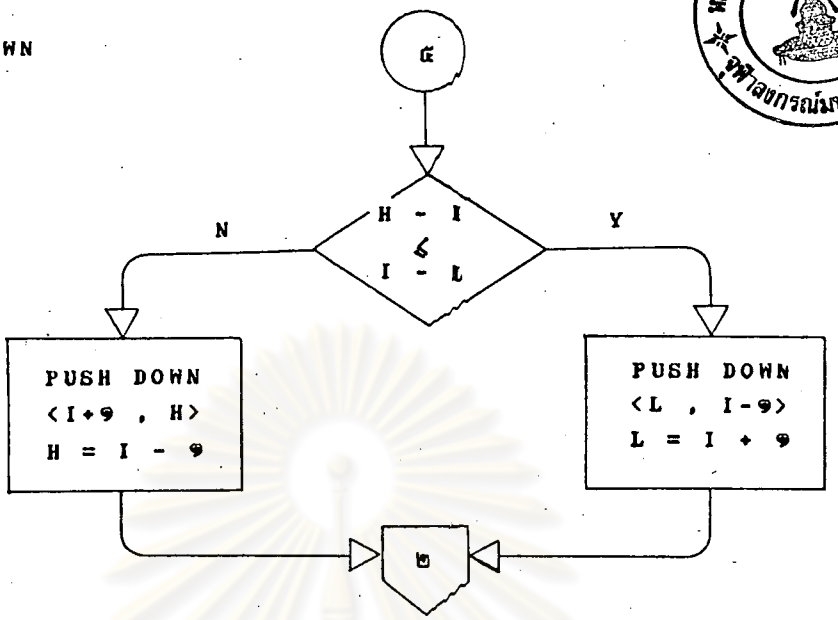


๔. LEFT TO RIGHT (GT)

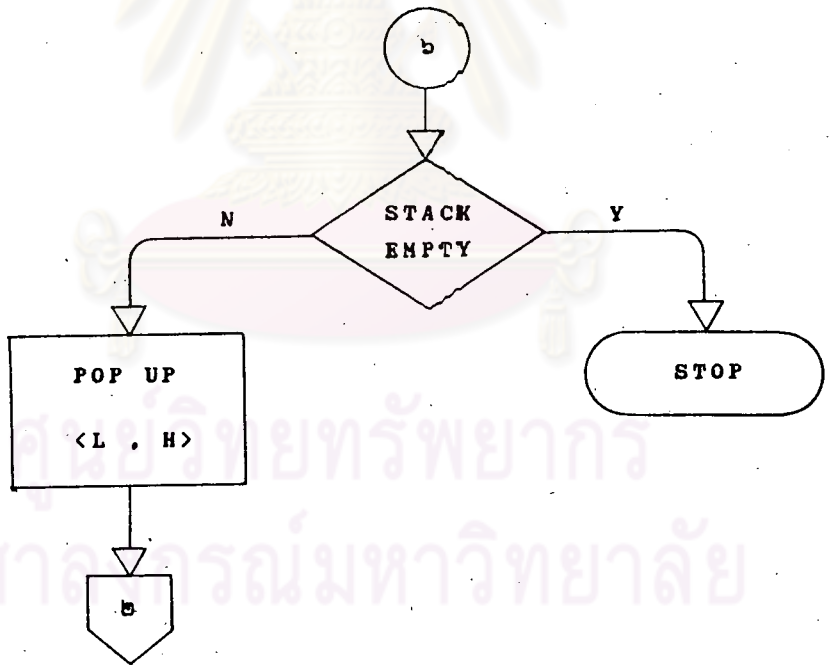


ผังงานที่ ๔.๒ แสดงการเรียงลำดับด้วยวิธี 'QUICK SORT' (ต่อ)

๕. PUSH DOWN



๖. POP UP



ผังงานที่ ๔.๒ แสดงการเรียงลำดับด้วยวิธี 'QUICK SORT' (ต่อ)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

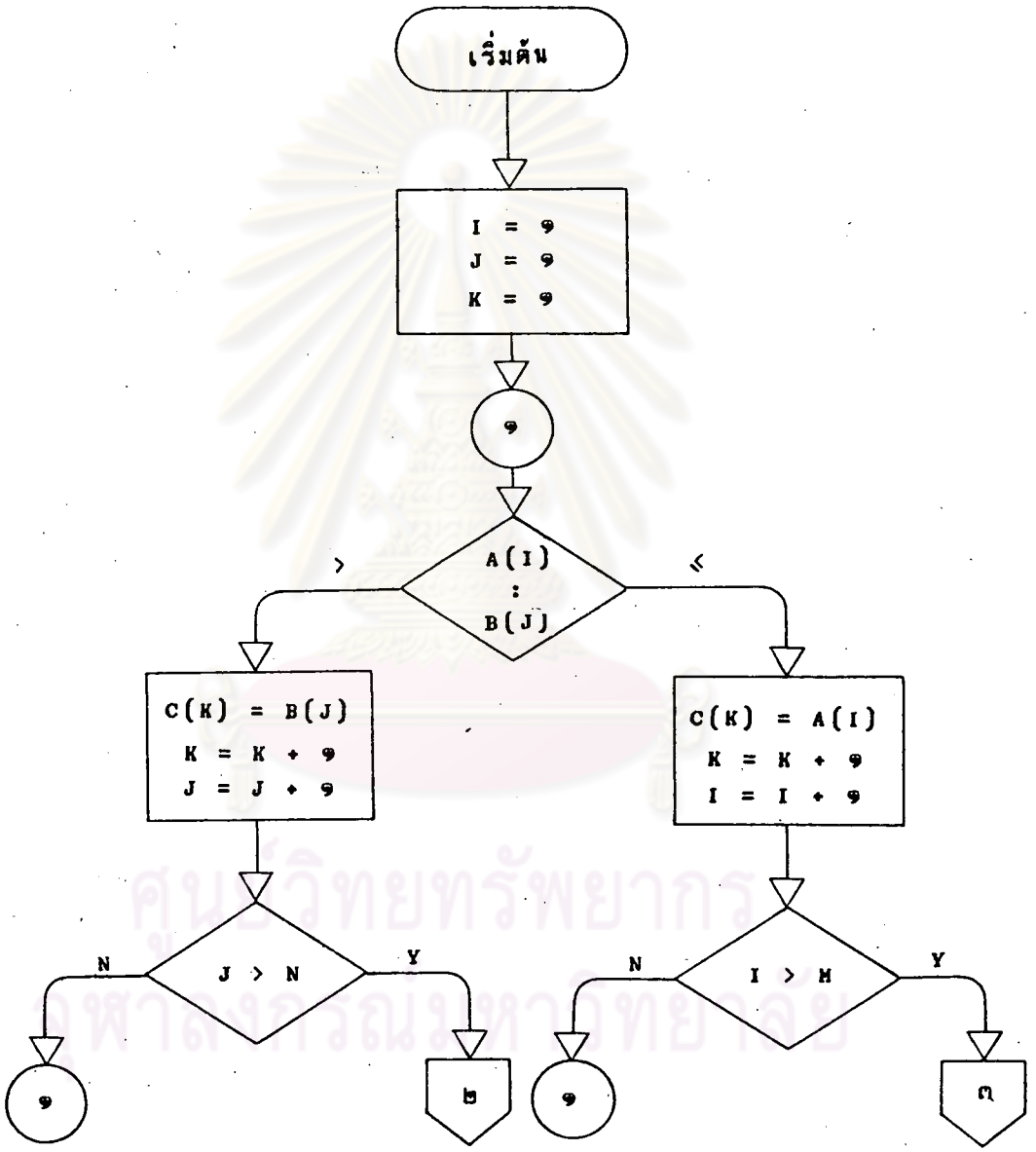
A(i) = คีย์ที่ i ในแฟ้มคีย์ที่ ๑

B(j) = คีย์ที่ j ในแฟ้มคีย์ที่ ๒

C = แฟ้มคีย์ที่ ๓ (ที่จะผสมได้)

M = จำนวนคีย์ในแฟ้มคีย์ที่ ๑

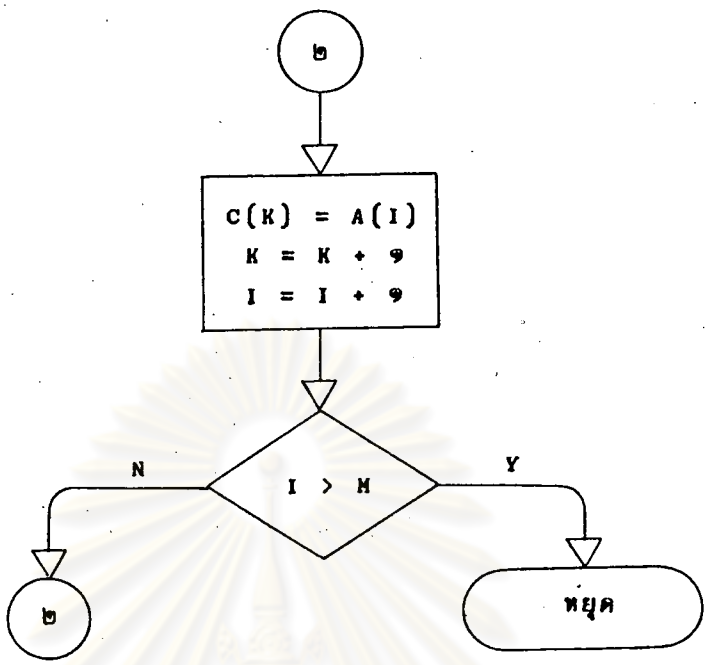
N = จำนวนคีย์ในแฟ้มคีย์ที่ ๒



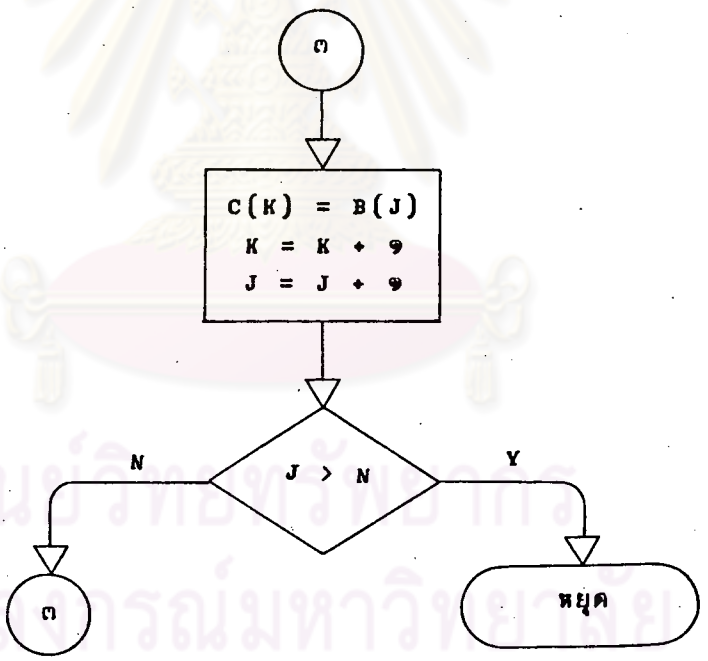
๑. A : B

ผังงานที่ ๔.๓ แสดงการผสมด้วยวิธี 'Simple Merge'

๒. EOFB



๓. EOFA



ผังงานที่ ๔.๓ แสดงการผสมด้วยวิธี 'Simple Merge' (ต่อ)