การใช้คอมพิวเตอร์ช่วยเก็บรักษาข้อมูล

ศรีศักดิ์ จามรมาน

คำนำ นักวิจัยส่วนมากในเมืองไทยขณะนี้มักจะมองข้ามขั้นตอนที่สำคัญไปขั้นตอนหนึ่ง นั้นคือการเก็บรักษาข้อมูล (Data Storage) ซึ่งในต่างประเทศมักจะใช้คอมพิวเตอร์เป็นเครื่องช่วย ฉะนั้นในบทความเรื่องนี้ผู้เขียนจึงขอเสนอเรื่องเกี่ยวกับ การจัดข้อมูลให้อยู่ในรูปแบบที่พร้อมจะเอาเข้าเครื่องคอมพิวเตอร์ได้ (Machine-Readable Form) เครื่องเตรียมและเก็บข้อมูล (Data Preparation Equipment and Media) การตรวจสอบแก้ไขข้อมูล (Data Editing and Purification) การจัดแฟ้มข้อมูล (File Organization) และระบบข้อมูลหลัก (Data Base System)

ข้อมูลที่เก็บหามาได้ใหม่ (Source Data) ข้อมูลที่เก็บหามาได้โดยวิธีต่างๆ เช่น จากการสำรวจและการสัมภาษณ์นั้น ถ้าเก็บหามาใหม่ๆ ยังไม่ได้เอาเข้าเครื่องคอมพิวเตอร์ ก็อาจจะเรียกได้ว่าเป็นข้อมูลใหม่ หรือ Source Data ซึ่งอาจจะอยู่ในหลายรูปแบบ เช่น ข้อความที่เขียนด้วยลายมือ ข้อความที่พิมพ์โดยเครื่องพิมพ์ดีด แบบฟอร์มที่กรอกด้วยลายมือ แบบฟอร์มที่กรอกโดยการพิมพ์ แบบฟอร์มที่ให้ใช้ดินสอขีดตามช่องที่กำหนดไว้ การ์ดคอมพิวเตอร์ที่ปรุรูไว้ให้เจาะรู

ถ้าข้อมูลอยู่ในแบบที่เป็นข้อความที่เขียนด้วยลายมือก็คงจะต้องมีพนักงานอ่านแล้วกรอกลงแบบฟอร์ม อาจจะด้วยลายมือหรือพิมพ์ดีดก็ได้

ถ้าข้อมูลอยู่ในแบบที่เป็นข้อความที่พิมพ์โดยเครื่องพิมพ์ดีด ก็อาจจะให้มีพนักงานอ่านแล้วกรอกลงแบบฟอร์ม หรืออาจจะใช้เครื่องอ่านชนิดที่เรียกว่า Optical Scanner

ถ้าข้อมูลอยู่ในแบบฟอร์ม ไม่ว่าจะกรอกด้วยมือ ด้วยการพิมพ์หรือโดยดินสอขีดตามช่อง ก็อาจจะส่งต่อไปเตรียมข้อมูลหรือใช้ Optical Scanner

ถ้าเป็นการ์ดคอมพิวเตอร์ก็ส่งเข้าเครื่องอ่านได้ทันที แต่ผู้เขียนเชื่อว่าในอนาคตอันใกล้นี้ วงการคอมพิวเตอร์ในเมืองไทยจะค่อย ๆ เลิกใช้การ์ดไปเรื่อย ๆ

การกรอกข้อมูลครั้งแรกนี้มีความสำคัญอย่างยิ่ง ถ้ามีการกรอกผิด ๆ ณ จุดนี้ การที่จะแก้ไขในขั้นตอนต่อไปก็อาจจะลำบาก ฉะนั้น นักวิเคราะห์ระบบ (System Analyst) หรือนักวิเคราะห์ข้อมูล (Information Analyst) ควรจะใช้เวลาจัดวางระบบแบบฟอร์มและการกรอกข้อมูลให้รัดกุม กรอกง่ายผิดน้อยและเสียเวลาน้อย ดังตัวอย่างต่อไปนี้

ออกแบบฟอร์มโดยกำหนดช่องไว้ให้กรอกเลขหรือตัวอักษรได้ช่องละหนึ่งตัวเท่านั้น คือ

A	B	1	2		

ออกแบบฟอร์มโดยพิมพ์คำตอบให้เลือกขีดช่องหน้าคำตอบ แทนที่จะเว้นช่องไว้ให้กรอก หรือถ้าเว้นที่ให้กรอกก็ควรมีคำตอบเป็นตัวเลขให้ เลือกมากรอกไม่ใช่ใช้ข้อความตามใจชอบของผู้กรอก

ออกแบบฟอร์มให้กรอกข้อมูลเพียงครั้งเดียว แต่มีสำเนาส่งให้หน่วยงานต่าง ๆ ใช้ได้โดยไม่ต้องกรอกใหม่

ออกแบบฟอร์มให้มีข้อมูลที่สอบยันกันได้เพื่อสะดวกแก่การ Edit และ Purify ซึ่งจะได้กล่าวถึงต่อไป

ในศูนย์คอมพิวเตอร์สมัยใหม่ที่มี Terminal เป็นจำนวนมาก ๆ Source data อาจจะได้โดยการให้คอมพิวเตอร์สัมภาษณ์ผู้มีข้อมูล เช่น ที่มหาวิทยาลัย Missouri มีการทดลองให้คนไข้สนทนากับคอมพิวเตอร์โดยตรง โดยใช้เครื่องพิมพ์ดีดหรือโทรทัศน์ให้คอมพิวเตอร์ถามปัญหาคนไข้ทีละปัญหา แล้วรอพิมพ์

คำตอบ หรือใช้ไฟฉายพิเศษส่องคำตอบที่ต้องการ ในวิธีแบบนี้เมื่อคอมพิวเตอร์ได้รับข้อมูลก็อาจจะจัดการ Edit และ Purify และเก็บเข้า Midia ตามที่ต้องการได้ทันที

เครื่องเตรียมข้อมูลและวัสดุเก็บข้อมูล (Data Preparation Equipments) เครื่องเตรียมข้อมูลที่ใช้กันอยู่ขณะนี้มีมากมายหลายอย่าง เช่น เครื่องเจาะบัตรและเครื่องทานบัตร (Key Punch and Verify) เครื่องเจาะแถบกระดาษ (Paper Tape Punch) เครื่องพิมพ์แถบแม่เหล็กในการ์ด (Magnetic Stripe Recorder) เครื่องพิมพ์แถบแม่เหล็กเป็นม้วน (Key-to-Tape) เครื่องพิมพ์แถบแม่เหล็กขนาดเล็ก (Key-to-Cassette or Minitape) เครื่องพิมพ์จานแม่เหล็กขนาดเล็ก (Diskett) เครื่องพิมพ์จานแม่เหล็ก (Key-to-Disk)

เครื่องเจาะบัตรเป็นเครื่องเตรียมข้อมูลที่ใช้กันแพร่หลายที่สุดในเมืองไทยขณะนี้ทั้ง ๆ ที่ควรจะเลิกใช้เป็นอย่างยิ่ง เพราะการใช้บัตรเป็นการสิ้นเปลืองโดยไม่จำเป็น เมื่อเจาะข้อมูลเข้าบัตรแล้วก็มักจะนำบัตรที่เจาะแล้วนั้นไปอ่านเข้าแถบแม่เหล็กหรือจานแม่เหล็ก เก็บแถบแม่เหล็กหรือจานแม่เหล็กนั้นไว้ใช้ต่อไปโดยไม่มีความจำเป็นจะต้องเก็บบัตรไว้อีก ในอเมริกานานมาแล้วเคยมีบริษัทประกันภัยใหญ่แห่งหนึ่งมีกฎว่าจะต้องเก็บบัตรข้อมูลของลูกค้าทั้งหมดไว้ มีปัญหามากมายต้องใช้โรงเก็บขนาดใหญ่ป้องกันอัคคีภัยลำบาก เมื่อเก็บไว้นานคุณภาพบัตรที่เจาะแล้วเสื่อม คอมพิวเตอร์อ่านลำบาก ในที่สุดก็ต้องตัดสินใจเลิกเก็บบัตร เก็บเป็นแถบแม่เหล็กหรือแม้แต่ Microfilm ก็ยังสะดวกกว่า

ปัญหาที่เกิดขึ้นบ่อย ๆ ในเมืองไทยเกี่ยวกับการใช้บัตรเก็บข้อมูลก็คือการกำหนดงบประมาณซึ่งมักจะไม่ค่อยใกล้เคียงความจริง ตัวอย่างเช่นหน่วยงานแห่งหนึ่งมีข้อมูลสามล้านชุด ก็คิดว่าควรใช้เงินประมาณหนึ่งล้านบาท นำข้อมูลนี้

เข้าคอมพิวเตอร์ ปรากฏว่าข้อมูลแต่ละชุดมีชื่อ ที่อยู่ อายุ เพศ และอื่น ๆ รวมประมาณชุดละสองร้อยตัวอักษร ฉะนั้น ถ้าจะใช้บัตรก็ต้องใช้ชุดละ 3 บัตร รวมเป็น 9 ล้านบัตร อัตราค่าเจาะและตรวจทานบัตรในเมืองไทยยังแตกต่างกันอยู่มาก คือจากบัตรละสี่สิบสตางค์ที่สถาบันบัณฑิตพัฒนบริหารศาสตร์ ถึงบัตรละสองบาทห้าสิบสตางค์ที่บริษัทคอมพิวเตอร์แห่งหนึ่ง ฉะนั้นค่าเจาะบัตรคงจะเป็นเงินจาก 3.6 ล้านบาท ถึง 22.5 ล้านบาท มากกว่าที่คาดไว้หลายเท่า

นอกจากปัญหาเรื่องงบประมาณแล้วก็มีปัญหาเรื่องเวลา จากตัวอย่างของข้อมูลสามล้านชุดนี้ เจ้าหน้าที่ที่เกี่ยวข้องก็คิดว่าคงจะนำเข้าคอมพิวเตอร์ได้เรียบร้อยภายในสามสี่เดือน สมมติว่ามาเจาะบัตรที่ สพบ. ซึ่งอาจจะเจาะและตรวจทานได้วันละประมาณสองพันบัตร ฉะนั้น รวมเก้าล้านบัตรจะต้องใช้เวลา 4,500 วัน มากกว่าสามสี่เดือนที่คาดไว้มากมาย ถ้าต้องการให้เสร็จเร็วขึ้นก็จะต้องเพิ่มจำนวนเครื่องเจาะบัตร เพิ่มจำนวนกะทำงาน หาพนักงานเจาะบัตรที่มีความสามารถมากขึ้น คือสำหรับราชการไทยหลายแห่ง กำหนดให้พนักงานเจาะบัตรได้วันละ 500–700 บัตร ของบริษัทฝรั่งเขากำหนดให้เจาะได้วันละ 1,500–3,000 บัตร ซึ่งก็อาจจะพิจารณาได้ว่าเป็นอัตราส่วนโดยตรงกับอัตราเงินเดือน คือพนักงานเจาะบัตรในราชการไทยอาจจะได้เงินเดือนเดือนละประมาณหนึ่งพันบาท แต่พนักงานของบริษัทฝรั่งในเมืองไทยได้เงินเดือนประมาณเดือนละสามพันบาท

นอกจากการใช้เครื่องเจาะบัตรแล้วในเมืองไทยก็มีการใช้เครื่องปรุแถบกระดาษอยู่เป็นจำนวนไม่น้อย เครื่องปรุแถบกระดาษนี้ก็คงจะค่อย ๆ เลิกใช้ไปเช่นเดียวกับเครื่องเจาะบัตร

ส่วนเครื่องพิมพ์แถบแม่เหล็กในการ์ดนั้นมักจะใช้กันในการทำบัญชีธนาคาร การค้าขายและการควบคุมสินค้าคงคลัง ทางด้านหน้าของการ์ดพิมพ์เป็นตัวอักษร

ตัวเลขให้อ่านด้วยตาเปล่าได้ ทางด้านหลังก็มีแถบแม่เหล็กให้เครื่องอ่านได้ วิธีนี้ก็ค่อย ๆ เลิกกันไปแล้วเช่นเดียวกับแถบกระดาษ

เครื่องที่ทันสมัยกว่าเครื่องทั้งสามชนิดที่กล่าวมาแล้วก็คือพวก Key-to-Tape และ Key-to-Disk ประโยชน์ที่ชัดเจนของเครื่องพวกนี้ก็คือตัดการใช้การ์ดออก ทำให้ประหยัดค่าใช้จ่ายลง ดังจะเห็นได้จากตัวอย่างเรื่องข้อมูลเก้าล้านการ์ดที่กล่าวมาแล้ว เฉพาะค่าการ์ดถ้าซื้อจำนวนมากใบละเจ็ดสตางค์ รวมเป็นเงินประมาณหกแสนบาท ทั้ง ๆ ที่ยังไม่คิดจำนวนการ์ดเผื่อเสีย ถ้าใช้แถบแม่เหล็กจะต้องจ่ายค่าแถบแม่เหล็กเพียงประมาณหนึ่งแสนบาท คือประมาณหนึ่งในหกของค่าการ์ด ประโยชน์อีกข้อหนึ่งของเครื่องเตรียมข้อมูลแบบใหม่ ๆ นี้ก็คือความสะดวกในการพิมพ์ ในการตรวจสอบแก้ไขข้อมูลที่พิมพ์เข้าไปแล้ว โดยเฉลี่ยแล้วเป็นที่ยอมรับกันว่าการใช้เครื่องเตรียมข้อมูลชนิดใหม่ ๆ นี้จะทำให้ประสิทธิภาพของพนักงานเตรียมข้อมูลดีขึ้นประมาณร้อยละยี่สิบ ฉะนั้น ถ้าเปลี่ยนจากเครื่องเจาะบัตรมาเป็นเครื่องแบบนี้ก็มักจะลดจำนวนเครื่องลงได้ด้วย

เครื่องเตรียมข้อมูลใหม่ ๆ นี้มีมากมายหลายชนิด แบ่งได้มากมายหลายแบบ เช่น แบ่งตามขนาดของแถบหรือจานแม่เหล็ก ถ้าแถบแม่เหล็กขนาดธรรมดาก็เรียก Key-to-Tape ถ้าแถบแม่เหล็กแบบ Cassette ที่ใช้อัดเพลงก็เรียก Key-to-Cassette ถ้าแถบแม่เหล็กเป็นม้วนเล็กเส้นผ่าศูนย์กลางม้วนประมาณ 2-3 นิ้ว ก็เรียกเป็น Key-to-Minitape ถ้าจานแม่เหล็กขนาดธรรมดาก็เรียก Key-to-Disk ถ้าจานขนาดเล็กก็เรียก Diskette เป็นต้น นอกจากจะแบ่งตามขนาดของแถบหรือจานแม่เหล็กแล้วก็อาจจะแบ่งตามความสามารถของเครื่องเป็น Intelligent Terminal และ Dummy Terminal

Intelligent Terminal ก็อาจจะแบ่งต่อไปเช่น Programmable Terminal และ Logic Terminal สำหรับ Programmable Terminal นั้น ผู้ใช้อาจจะเขียน

โปรแกรมให้ Terminal ทำตาม เช่นเดียวกับเขียนโปรแกรมให้คอมพิวเตอร์ทำตาม นั่นคือ Terminal จะทำหน้าที่เป็น Front-end Minicomputer ใช้ทำได้ทั้งพิมพ์ ตรวจทาน (Verify) และตรวจสอบแก้ไข (Edit and Purify) ไปในตัวเลย ทำให้ประหยัดเวลาคอมพิวเตอร์ไปได้อีกต่อหนึ่งด้วย เครื่อง Programmable Terminal นั้น มี Storage ของตัวเองเก็บข้อมูลไว้สำหรับใช้กับคอมพิวเตอร์ภายหลังได้ ส่วน Logic Terminal ก็ตรวจสอบข้อมูลได้แต่ไม่มี Storage สำหรับเก็บข้อมูลต้องส่งเข้าคอมพิวเตอร์ไปประมวลผลหรือให้คอมพิวเตอร์ส่งต่อเข้าแถบหรือจานแม่เหล็ก

สำหรับ Dummy Terminal นั้น อาจจะเป็นพิมพ์ดีดหรือโทรทัศน์ที่ต่อเข้ากับคอมพิวเตอร์ โปรแกรมในตัวเองก็ไม่มี Logic ในตัวเองก็ไม่มี ทำได้เพียงส่งข้อมูลไปให้คอมพิวเตอร์ ให้คอมพิวเตอร์จัดตรวจสอบส่งสัญญาณมาออกทาง Dummy Terminal ให้ผู้ใช้จัดแก้ไขโดยพิมพ์การแก้ไขที่พิมพ์ดีดส่งไปให้เครื่องคอมพิวเตอร์แก้ไขให้

การระบุชนิดของ Terminal ที่จะใช้นี้ควรใช้ภาษาอังกฤษกำกับ มิฉะนั้นจะเกิดการเข้าใจผิดได้ง่าย ดังตัวอย่างเช่น หน่วยราชการของไทยแห่งหนึ่งเมื่อเร็ว ๆ นี้ต้องการจะใช้ Intelligent Terminal แต่ฝ่ายกฎหมายไม่ยอมใช้ภาษาอังกฤษ จึงแปลเป็นภาษาไทยว่าเครื่องเตรียมข้อมูลที่สามารถปฏิบัติงานได้โดยอิสระ เมื่อประกวดราคากันก็ต้องยอมรับ Dummy Terminal ซึ่งราคาต่ำกว่า Intelligent Terminal โดยผู้ขาย Dummy Terminal อ้างว่าเครื่องของเขาปฏิบัติงานได้โดยอิสระ คือจะใช้เครื่องเดียวก็ได้ หลาย ๆ เครื่องก็ได้ แต่ละเครื่องไม่ขึ้นแก่กันและกัน แต่ตามความจริงแล้วเครื่อง Intelligent Terminal ทำงานได้มากกว่า และสามารถปฏิบัติงานได้แบบ Off-Line คือปฏิบัติงานโดยตัวของ

มันเอง ไม่ต้องเชื่อมโยงกับเครื่องคอมพิวเตอร์ ส่วนเครื่อง Dummy Terminal นั้น ถ้าเครื่องคอมพิวเตอร์หยุดทำงานเมื่อใด มันก็ต้องหยุดด้วย

การตรวจสอบแก้ไขข้อมูล (Data Editing and Purification) ข้อมูลที่นำเข้าคอมพิวเตอร์ในครั้งแรกนั้น มักจะมีที่ผิดอย่างชนิดที่เรียกว่า Bias และ Error ตัวอย่างของ Bias ก็มีเช่น เวลาผู้จัดการถามพนักงานขายว่าจะขายสินค้าได้เท่าใดในช่วงระยะเวลาใดเวลาหนึ่งในอนาคต พนักงานขายมักจะให้ตัวเลขที่สูงกว่าความเป็นจริง หรือเมื่อลูกค้านำเครื่องมือเครื่องใช้มาซ่อม ผู้รับซ่อมมักจะแจ้งกำหนดเวลาที่จะซ่อมเสร็จเรียบร้อยน้อยกว่าที่จะเป็นจริง ถ้าแจ้งว่า 3 วันเสร็จก็มักจะเป็น 7 วันเป็นต้น ฉะนั้น Bias นี้ เราอาจจะให้คอมพิวเตอร์จดจำไว้ว่าผู้มี Bias มักจะ Overestimate หรือ Underestimate ประมาณเท่าใดแล้ว Calibrate ข้อมูลให้ใกล้ความเป็นจริงมากขึ้น

ส่วนความผิดพลาดชนิดที่มิได้มาจาก Bias คือ ชนิดที่เป็น Error ก็อาจจะเกิดขึ้นเพราะเหตุต่าง ๆ ต่อไปนี้ การนับผิด ชั่งตวงวัดผิดในการเก็บหาข้อมูล ในกรณีที่ Source Data ของงานหนึ่งได้มาจากการประมวลข้อมูลจากอีกงานหนึ่ง Source Data ชนิดนี้ก็อาจจะผิดเพราะการประมวลข้อมูลผิด ข้อมูลสูญหาย ลอกข้อมูลผิดหรือแก้ผิด ใช้ตารางผิดในการหาข้อมูลเพิ่มเติม เช่น มีชื่อคนแล้วต้องการที่อยู่ แต่ใช้ตารางที่อยู่โบราณ ก็ได้ที่อยู่โบราณ จงใจทำให้ผิด

การตรวจสอบแก้ไขข้อมูลอาจจะทำได้หลายระดับ เช่น เริ่มโดยมีบรรณาธิกร อ่านแบบฟอร์มที่ส่งเข้ามาว่ามีตัวใดอ่านไม่ออก หรือนอกลู่นอกทางหรือไม่ ถ้ามีก็อาจจะจัดการสอบถามผู้กรอก จัดแก้ไขให้เรียบร้อยก่อน ต่อจากนั้นเมื่อจัดพิมพ์เข้าเครื่องก็อาจจะมีการ Verify คือให้คนสองคนพิมพ์ข้อมูลเดียวกันโดยใช้หลักที่ว่าถ้าคนสองคนพิมพ์ผิดก็ไม่ควรจะผิดที่เดียวกัน ต่อจากการ

Verify แล้วก็อาจจะมีการตรวจสอบแก้ไขโดยโปรแกรม ซึ่งอาจจะตรวจสอบจุดต่าง ๆ ดังต่อไปนี้

Identification Code เช่น เลขประจำตัว เลขหมายเครื่อง เลขที่บัญชี เลขหมายสินค้า เป็นต้น ในกรณีนี้อาจจะใช้ Check Digit คือถ้า ID มี 3 ตัว ก็เพิ่มเลขให้อีกตัวเป็นสี่ตัว โดยการคำนวณที่ตรวจสอบได้ เช่น อาจจะใช้ **Weight** 731

$$W_1 \quad W_2 \quad W_3$$

$$I_1 \quad I_2 \quad I_3 \quad C$$

คือให้ C เป็นเลขตัวสุดท้ายของ $W_1 I_1 + W_2 I_2 + W_3 I_3$

สมมติว่า ID เป็น 245 ก็จะได้ C เป็นเลขตัวท้ายของ $7 \times 2 + 3 \times 4 + 1 \times 5$ คือ 1 นั่นคือใช้ ID ใหม่เป็น 2451 ถ้ามีการพิมพ์ผิดขึ้น เช่น พิมพ์เป็น 2541 เครื่องก็จะบอกได้ทันทีว่าผิดเพราะ $1 \neq$ ตัวสุดท้ายของ $7 \times 2 + 3 \times 5 + 1 \times 4$

Transaction Code คือถ้าใช้รหัสระบุว่าข้อมูลเป็นข้อมูลชนิดใด ก็จะต้องตกลงกันล่วงหน้าว่ามีรหัสอะไรบ้าง เครื่องจะได้ตรวจสอบว่ารหัสที่ใช้นั้นถูกต้องหรือไม่ ถ้ารหัสแบบใดจะมีการสอบยันได้อย่างไร

Character คือต้องตกลงกันไว้ล่วงหน้าว่าข้อมูลชนิดใดจะใช้ตัวเลขตัวอักษรอย่างใด เช่น อายุ ต้องเป็นตัวเลข ชื่อต้องเป็นตัวอักษร ถ้าพิมพ์อายุเป็น IJ เครื่องจะได้บอกได้ทันทีว่าผิด บางเรื่องเช่น เพศอาจจะใช้เลขเป็น 1–ชาย และ 2–หญิง หรือ M –ชาย และ F –หญิง ควรจะใช้เลขหมดหรืออักษรหมด ไม่ควรให้สลับกันได้เพราะจะผิดพลาดได้ง่าย

Field Length, Sign and Composition ถ้าข้อมูลชนิดใดมีความยาวจำกัด มีเครื่องหมายมีส่วนประกอบเป็นรูปแบบใด เช่น สองตัวแรกเป็นอักษร สามตัวหลังเป็นเลข ก็ต้องจัดให้เครื่องตรวจสอบด้วย

Missing Data ต้องตรวจสอบดูว่าข้อมูลมีครบตามที่ตกลงกันไว้หรือไม่ เช่น ถ้าส่วนหนึ่งระบุว่ามีบุตร ต้องให้ระบุชื่อบุตร เป็นต้น

Sequence Test ตรวจสอบว่าข้อมูลอยู่ใน Sequence ที่ควรจะเป็นหรือไม่ เช่น วันที่ จากก่อนไปหลัง เป็นต้น

Limit, Range or Reasonableness Test ต้องตรวจสอบว่าข้อมูลอยู่ในขนาดที่ควรจะเป็นหรือไม่ เช่น อายุ ไม่ควรเกิน 130 หรืออายุข้าราชการควรอยู่ในขนาด 18-60 ราคารวมของสินค้าต้องเท่ากับผลรวมของราคาแต่ละรายการ เป็นต้น

การตรวจสอบแก้ไขข้อมูลนี้ ควรจะจัดทำในตอนต้น ๆ ของกระบวนการ เช่น ถ้าเป็นแบบ Batch Processing ก็ควรทำก่อนหรือหลังการ Sort ครั้งแรก ถ้าเป็นแบบ On-Line หรือ Intelligent Terminal ก็ทำทันทีที่พิมพ์เข้า

เมื่อพบ Error เมื่อใด ควรจะให้เครื่องจดไว้ว่ามี Error แล้วแยกข้อมูลชุดนั้นออกจากข้อมูลทั้งหมด ส่งข้อมูลชุดที่ผิดไปให้ผู้กรอกข้อมูลจัดการแก้ไข ให้เครื่องพิมพ์ Aging ให้ว่าข้อมูลใดส่งไปนานเท่าใดแล้วยังมิได้แก้ เพื่อสะดวกแก่การติดตาม

โดยทั่ว ๆ ไปแล้ว ตกลงกันว่าถ้าข้อมูลในงานใดผิดมากกว่าร้อยละห้า ให้ถือว่าจะต้องพิจารณาแก้ไขปรับปรุงระบบใหม่ได้ อาจจะเป็นเพราะแบบฟอร์มไม่ดี ระเบียบวิธีกรอกแบบฟอร์มไม่ดี ผู้เตรียมข้อมูลไม่มีความสามารถเพียงพอก็ได้ ต้องจัดการแก้ไขให้ถูกจุด

การจัดแฟ้มข้อมูลและระบบข้อมูลหลัก นอกจากการตรวจสอบแก้ไขข้อมูลแล้ว เรื่องสำคัญที่ต้องพิจารณาอีกเรื่องหนึ่งก็คือจะจัดแฟ้มข้อมูลอย่างไร ในเรื่องนี้มีศัพท์ทางคอมพิวเตอร์ที่นักวิจัยคิดจะใช้คอมพิวเตอร์ควรทราบมากมายหลายอย่าง ดังตัวอย่างต่อไปนี้

Bit คือ ข้อมูลที่มีได้เพียงสองค่า ศูนย์หรือหนึ่ง อาจจะเป็น True กับ False หรือ Yes กับ No หรือ ขาวกับดำก็ได้

Character คือ ข้อมูลที่มีค่าเป็นตัวอักษร เช่น A,B,C ตัวเลข เช่น 0,1,2 หรือเครื่องหมายพิเศษ เช่น +, −, ×

Field หรือ Item คือข้อมูลที่มีตั้งแต่หนึ่ง Character ขึ้นไป เช่น ชื่อ บ้านเลขที่ อายุ เป็นต้น

Record คือข้อมูลที่มีตั้งแต่หนึ่ง Field ขึ้นไป เช่น Personnel Record, Inventory Record เป็นต้น

File คือข้อมูลหลาย ๆ Record รวมกันเช่น Personnel File ก็อาจมี Record ของพนักงานทุกคน เป็นต้น

ขนาดของ Field และขนาดของ Record อาจจะคงที่ (Fixed) หรือไม่แน่นอน (Variable) ตัวอย่างเช่น ชื่อคนอาจจะกำหนดให้มีขนาดคงที่เป็น 30 characters หมด ชื่อใดสั้นก็เติม Blank เข้าไปให้ครบสามสิบ หรืออาจจะจัดให้เป็น Variable คือชื่อใครยาวเพียงใดก็ใช้ field ขนาดนั้น ถ้าใช้ field ขนาดไม่คงที่ก็จะต้องหาวิธีทำให้เครื่องทราบว่าเมื่อใดจะเป็นขนาดเท่าใด มีวิธีที่ใช้ ๆ กันอยู่สองวิธีคือ

ระบุความยาว โดยตกลงกันว่า character แรกสามตัวระบุความยาวของ field หรืออะไรทำนองนั้น

ใช้เครื่องหมายพิเศษระบุที่จบ เช่น เติมเครื่องหมาย ";" ที่ท้าย field ทุก field

ในกรณีของ Record ถ้ามี Record มาก ๆ เป็น หมื่น ๆ แสน ๆ การใช้ fixed length อาจจะเปลืองที่โดยไม่จำเป็น เช่น ถ้าขีดความยาวคงที่ไว้ 200

แล้วโดยเฉลี่ยใช้เพียง 100 ถ้ามีหนึ่งหมื่น Record ก็เปลืองที่ไปหนึ่งล้านที่โดยไม่จำเป็น ถ้าจัดแบบ Variable Length เวลาจะเรียกมาดูมาใช้ก็ไม่สะดวก เพราะต้องเสียเวลาหาจุดจบของแต่ละ Record ก่อน ฉะนั้นทางออกที่ง่าย ๆ ก็คือจัดให้แต่ละ Record มีสองส่วน ส่วนหนึ่งที่ต้องการเรียกมาใช้บ่อย ๆ เช่น Fixed Length แล้วมีเครื่องหมายพิเศษบอกทางต่อไปยังอีกส่วนหนึ่งที่เป็น Variable Length

ชนิดของแฟ้มข้อมูลก็อาจจะแบ่งออกหลายอย่างหลายวิธี เช่น แบ่งโดยการใช้และแบ่งโดยการเก็บ ถ้าแบ่งโดยการใช้ก็มีชนิดที่สำคัญ ๆ อยู่ 4 ชนิด คือ

Master File คือแฟ้มข้อมูลที่ค่อนข้างจะถาวร ใช้อ้างอิง เช่น รายชื่อบุคคลากรทั้งหมด รายชื่อวัสดุอุปกรณ์ทั้งหมด เป็นต้น

Transaction File หรือ Detail File คือแฟ้มของรายการที่เกิดขึ้นใหม่เป็นระยะ ๆ ใช้สำหรับมาปรับปรุงแก้ไข Master File เช่น แฟ้มรายการผู้ย้ายที่อยู่ในเดือนนี้

Report File เป็นแฟ้มที่เก็บข้อมูลที่แยกมาจาก Master File เพื่อจะใช้ทำรายงาน

Sort File คือแฟ้มชั่วคราวที่ลอกมาจากแฟ้มชนิดใดชนิดหนึ่งในสามชนิดแรก เพื่อจะจัดลำดับใหม่ โดยหลักการทางคอมพิวเตอร์แล้ว การจะจัดลำดับแฟ้มใดแฟ้มหนึ่งควรจะลอกมาเข้า Sort File เผื่อเกิดการเสียหายขึ้นระหว่างการจัดลำดับ จะได้มีแฟ้มเก่าเอากลับมาใช้ได้

ถ้าแยกประเภทแฟ้มข้อมูลตามวิธีเก็บภายในก็มีประเภทที่ใช้กันอยู่บ่อย ๆ คือ Sequential, Indexed และ Indexed Sequential

ใน Sequential File เราจัดข้อมูลเรียงลำดับกันไปตามลำดับที่ส่งข้อมูลเข้าเครื่อง การจัดแบบนี้อาจจะใช้การ์ดหรือแถบแม่เหล็กหรือจานแม่เหล็กก็ได้ การ

จะเรียก Record ใด record หนึ่งมาดูก็ต้องเริ่มเรียกตั้งแต่ Record ที่หนึ่งไป สองไปสามเรื่อย ๆ ไปจนกว่าจะได้ Record ที่ต้องการ จะเรียก Record ใด record หนึ่งมาดูโดยตรงโดยไม่ดู Record ตอนต้นก่อนมิได้ ข้อดีที่สำคัญของการใช้ Sequential File ก็คือเสียค่าใช้จ่ายน้อย เพราะใช้แถบแม่เหล็กได้ ไม่จำเป็นต้องใช้จานแม่เหล็ก ซึ่งราคาสูงกว่าแถบแม่เหล็กเป็นสิบ ๆ เท่า แต่การใช้แถบแม่เหล็กก็มีข้อเสียคือเครื่องอ่านแถบแม่เหล็กได้ช้ากว่าจานแม่เหล็กเป็นอัตราส่วนมากกว่าหนึ่งต่อสิบ ถ้าเรามีทางเลือกได้ คือถ้าเครื่องที่เราจะใช้มีทั้งแถบแม่เหล็กและจานแม่เหล็กเราควรจะเลือกใช้ Sequential File เมื่อ Activity Rate ของ File อยู่ในเกณฑ์สูง คือ เราจำเป็นจะต้องเรียก Record ใน File มาดูมาแก้เป็นจำนวนมาก ถ้า Activity Rate ของ file ต่ำ เช่น ถ้ามีหนึ่งหมื่น records แล้วจะเรียกมาพิจารณาครั้งละเพียงสิบ records ก็ไม่ควรใช้ Sequential File

ใน Indexed หรือ Random หรือ Direct Access File นั้น ถ้าเราต้องการจะพิจารณา Record ใด เครื่องก็จะนำ Record นั้นมาให้เราโดยตรงโดยมิต้องอ่าน Record ที่มาก่อนหน้า Record นั้น การจัดแบบนี้ใช้แถบแม่เหล็กไม่ได้ จะต้องใช้จานแม่เหล็ก และจะต้องตกลงกันเลือก Filed หนึ่งใน Record เป็น Key เวลาจะเรียก Record ใดก็ต้องเรียกโดย Key การที่เราจะเรียก Record โดยใช้ Key ได้นี้ เครื่องก็จะต้องจัดทำตารางหรือ Directory เอาไว้เองว่า Record ใดอยู่ที่ใด ในระบบเล็ก ๆ อาจจะจัด Directory ได้เพียงตัวเดียวระดับเดียว ในระบบใหญ่ ๆ อาจจะจัดได้หลายตัวหลายระดับ ตัวอย่างของการจัด Key ได้หลายตัวก็คือ ถ้าเป็นแฟ้มข้อมูลแบบสมุดโทรศัพท์ก็อาจจะจัดให้มี Key ได้หลายตัว เช่น ชื่อ หมายเลขโทรศัพท์ หรือบ้านเลขที่ เป็นต้น ตัวอย่างของการจัด Key หลายระดับก็เช่นแบบที่เรียกว่า Bucket คือจัด Record หลาย ๆ Record รวมกันเป็น Bucket แล้วใน Bucket ก็ใช้การหาแบบ Sequential

ตามที่ได้กล่าวแล้ว การเลือกใช้ Indexed File นั้น ควรจะทำเมื่อ Activity Rate ของ File ต่ำ เช่น ในระบบ Inquiry การใช้ File แต่ละครั้งอาจจะหา Record เพียง Record เดียว เช่น จากหนึ่งหมื่น Record ถ้าจะขอดูเพียง Record เดียว ถ้าใช้ Sequential File ก็จะต้องอ่านจาก Record ที่หนึ่ง จนพบ นั่นคือโดยเฉลี่ยแล้วจะต้องดูถึงห้าพัน Record ถ้าใช้ Indexed File ก็ขอดูโดย ตรงได้เลย

ข้อเสียของ Indexed File ก็คือ ถ้าเผอิญอยากจะพิมพ์ข้อมูลทั้งหมดใน แฟ้มนั้นออกมาให้เรียงตามลำดับอะไร เช่น ลำดับชื่อ ก็จะต้องไปดู Directory ว่าชื่อนี้อยู่ที่ใด ไปเอามาพิมพ์แล้วไปดู Directory หาที่อยู่ของชื่อต่อไป ถ้าเป็น Sequential File ที่เรียงตามชื่ออยู่แล้วก็เริ่มตั้งแต่ต้นไปจนจบ อ่านแล้วพิมพ์ เรียงกันออกมาโดยไม่มี Directory ให้ต้องไปดู

ฉะนั้นการประนีประนอมระหว่าง Sequential File กับ Indexed File ก็คือ ใช้ Indexed Sequential File นั่นคือจัดเก็บข้อมูลในจานแม่เหล็กเรียงกันตาม Key แล้ว มี Directory บอกว่า Key ใดอยู่ที่ใดด้วย ฉะนั้นจะพิมพ์เรียงตาม Key ก็สะดวก จะเรียก Record โดยมี Key ใดหนึ่งก็ไม่เสียเวลามาก การเก็บ ข้อมูลส่วนมากในสมัยนี้มักจะใช้ Indexed Sequential แต่ข้อเสียก็ต้องมีเป็น ธรรมดา ที่สำคัญก็คือเปลืองที่เปลืองเวลาเมื่อต้องการจะเพิ่มเติมข้อมูลเข้าแฟ้ม เพราะจะต้องจัดย้ายข้อมูลอื่น ๆ ออกให้มีที่ไว้ข้อมูลใหม่ตรงกลาง ๆ ได้ และต้อง จัด Index ใหม่

เมื่อกล่าวถึงการจัดแฟ้มข้อมูลกันมาเพียงนี้แล้ว ก็ควรจะต้องกล่าวถึงหลัก การใหม่ ซึ่งเรียกว่าระบบข้อมูลหลัก (Data Base System) ซึ่งใครๆ ในวงการ คอมพิวเตอร์ในเมืองไทยก็อยากจะใช้กันทั้งนั้น หลักการสำคัญของระบบข้อมูล

หลักก็คือ แทนที่จะมีแฟ้มข้อมูลทั้งเล็กทั้งใหญ่แยกกันมากมายหลายแฟ้ม ก็จัดให้มีแฟ้มรวมเพียงแฟ้มเดียว เช่น ในมหาวิทยาลัย แทนที่จะมีแฟ้มหนึ่งเรื่องเงินเดือน (Payroll) อีกแฟ้มหนึ่งเรื่องคุณวุฒิ (Skill Inventory) อีกแฟ้มหนึ่งเรื่องการยืมหนังสือห้องสมุด และอีกแฟ้มหนึ่งเรื่องการสอน รวมทั้งสี่แฟ้มนี้มีชื่ออาจารย์อยู่หมดรวมสี่ครั้ง เปลืองที่โดยไม่จำเป็น จึงน่าจะรวมเป็นแฟ้มเดียว มีข้อมูลทุกอย่างเกี่ยวกับอาจารย์แต่ละคน ชื่ออาจารย์คนใดคนหนึ่งก็อยู่ในคอมพิวเตอร์เพียงครั้งเดียว เวลาจะแก้ไขข้อมูล เช่นอาจารย์หญิง เปลี่ยนนามสกุล ก็แก้เพียงครั้งเดียวไม่ต้องตามแก้ทุกแฟ้ม

อย่างไรก็ตามระบบข้อมูลหลักนี้ จะต้องมีการควบคุมบำรุงรักษาเป็นอย่างดี จะต้องตั้งตำแหน่งผู้รักษาระบบข้อมูลหลัก เป็นเจ้าหน้าที่ระดับสูง ใครจะเปลี่ยนแปลงแก้ไข จะใช้ข้อมูลอย่างไร เมื่อไร ต้องได้รับความเห็นชอบจากผู้รักษาระบบข้อมูลหลัก

นอกจากตัวคนที่จะเป็นผู้รักษาระบบข้อมูลหลักแล้ว ก็จะต้องมีโปรแกรมเป็น Data Base Management System สำหรับสร้างแฟ้ม (Create) ปรับปรุงแฟ้ม (Update) เลือก (Select) หาข้อมูล (Retrieve) จัดลำดับ (Sort) และจัดทำงาน (Report generation) โปรแกรมชนิดนี้เป็นโปรแกรมใหญ่ ใช้ส่วนความจำเป็นจำนวนมาก จะเขียนเองก็คงต้องใช้เวลาเป็นปี จึงมักนิยมเช่า หรือซื้อจากบริษัทต่างๆ เช่น ระบบ MIS และ DL/1 จาก IBM ระบบ Mark IV จาก Informatics Inc. ระบบ TOTAL จาก Sincom System, Inc. และระบบ System 2000 จาก MRI Systems Corp.

สรุป จากบทความเรื่องนี้ท่านผู้อ่านคงจะได้ทราบคร่าว ๆ ว่า เมื่อมีข้อมูลที่จะต้องเก็บไว้ใช้ในการวิจัยเป็นจำนวนมากต้องใช้เครื่องคอมพิวเตอร์ช่วยเก็บ ก็มีเรื่องที่จะต้องพิจารณาหลายเรื่อง คือการจัดข้อมูลให้อยู่ในรูปแบบที่คอมพิวเตอร์อ่านได้ การเตรียมและเก็บข้อมูล การตรวจสอบแก้ไขข้อมูล การจัดแฟ้มข้อมูลและระบบข้อมูลหลัก ฉะนั้นในการวางแผนการวิจัยที่ต้องใช้ข้อมูลมาก ๆ ต้องใช้คอมพิวเตอร์ ผู้วางแผนควรจะได้ศึกษาหาความรู้เกี่ยวกับการใช้คอมพิวเตอร์ในด้านนี้ให้เพียงพอ หรือถ้าไม่มีเวลาก็ควรจะปรึกษากับผู้เล่นผู้รู้ทางด้านคอมพิวเตอร์ เพื่อจะได้จัดเตรียมเวลา เตรียมงบประมาณให้ถูกต้องในระยะที่วางแผนและปฏิบัติตามขั้นตอนตามวิธีการที่ถูกต้อง ให้ได้ข้อมูลที่ดี ที่บริสุทธิ์ เก็บรักษาไว้ในรูปแบบที่เหมาะสมแก่การใช้สำหรับการวิจัยต่อไป