บทที่ 2

แนวคิดและทฤษฎี

2.1 โปรแกรมซีดีเอส/ไอซีส

CDS/ISIS ย่อมาจาก Computerized Documentation System/Integrated Set of Information Syatems เป็นโปรแกรมสำเร็จรูปที่พัฒนาบนเครื่องคอมพิวเตอร์ขนาดใหญ่โดยองค์การยูเนสโก เมื่อปี ค.ศ.1975 สำหรับประมวลผลและจัดการระบบฐานข้อมูลตัวอักษร ที่ไม่จำกัดขนาด (unlimited textual /non-numeric database) และเผยแพร่ให้ประเทศต่างๆ นำไปใช้ได้ เพื่อประโยชน์ทางการศึกษา เฉพาะกับหน่วยงาน องค์การ สถาบัน ที่ประกอบกิจการโดยไม่หวังผลกำไร (non-profit institution)

ต่อมาในปี ค.ศ.1985 ยูเนสโกได้พัฒนาโปรแกรมดังกล่าวบนเครื่องไมโครคอมพิวเตอร์เป็นครั้งแรก เรียกว่า มินิ-ไมโคร ซีดีเอส/ไอซีส พัฒนาด้วยภาษาปาสคาล ให้ใช้กับเครื่องคอมพิวเตอร์ ขนาด 16 บิต ขึ้นไป ที่ใช้ระบบปฏิบัติการ MS-DOS มีหน่วยความจำหลักไม่น้อยกว่า 256 กิโลไบต์ และมีจานแม่เหล็กที่สามารถจุข้อมูลได้ไม่น้อยกว่า 10 เมกาไบต์

ยูเนสโกได้พัฒนารุ่นที่สองของโปรแกรมนี้ได้เป็นผลสำเร็จสมบูรณ์ เมื่อเดือนมีนาคม ค.ศ.1989 และเริ่มเผยแพร่ให้แก่ National Distributor ของประเทศสมาชิกของยูเนสโกตั้งแต่นั้นมา

ในประเทศไทยมีการทดลองใช้โปรแกรมนี้อย่างแพร่หลายทั้งสถาบันการศึกษา องค์การ และหน่วยงานภาครัฐบาล รัฐวิสาหกิจ และธนาคาร จุฬาลงกรณ์มหาวิทยาลัยได้เริ่มติดตั้งโปรแกรมซีดีเอส/ไอซีส บนเครื่องคอมพิวเตอร์ขนาดใหญ่ ณ สถาบันบริการคอมพิวเตอร์ ตั้งแต่เดือนตุลาคม ค.ศ.1983 เพื่อรองรับโครงการจัดทำรวมรายชื่อวารสารในประเทศไทยด้วยคอมพิวเตอร์ ฉบับพิมพ์ครั้งที่ 2 ซึ่งสถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย รับผิดชอบเป็นผู้ประสานงานโครงการขณะนั้น

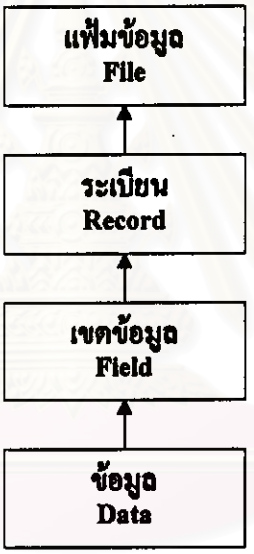
สถาบันวิทยบริการ จุฬาฯในฐานะ National Distributing Center ของโปรแกรมมินิ-ไมโคร ซีดีเอส/ไอซีส ในประเทศไทย ได้ทดลองใช้โปรแกรมบนเครื่องไมโครคอมพิวเตอร์ ตั้งแต่เดือนมกราคม ค.ศ. 1986 เป็นต้นมา และได้พัฒนาโปรแกรมให้สามารถรับข้อมูลภาษาไทย ด้วยรหัสภาษาไทยของมหาวิทยาลัยเกษตรศาสตร์ และรหัสภาษาไทยของสำนักงานมาตรฐานอุตสาหกรรม (สมอ.) ได้เป็นผลสำเร็จ เมื่อเดือนมิถุนายน ค.ศ.1988 สำหรับรุ่น 2.3 สถาบันวิทยบริการ ได้รับโปรแกรมพร้อมคู่มือฉบับสมบูรณ์จากยูเนสโก เมื่อเดือนมิถุนายน ค.ศ.1989 และพัฒนาให้รับข้อมูลภาษาไทยได้ด้วยรหัสภาษาไทยทั้งของมหาวิทยาลัยเกษตรศาสตร์ และสำนักงานมาตรฐานอุตสาหกรรม โดยเผยแพร่ในชื่อ มินิ-ไมโคร ซีดีเอส/ไอซีสไทย รุ่น 2.3

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สำหรับในวิทยานิพนธ์ฉบับนี้ จะเรียกชื่อโปรแกรมมินิ-ไมโคร ซีดีเอส/ไอซีส หรือโปรแกรม MINI-MICRO CDS/ISIS ว่า โปรแกรมซีดีเอส/ไอซีส

2.2 ฐานข้อมูลซีดีเอส/ไอซีส

ฐานข้อมูลซีดีเอส/ไอซีส (จุฬาลงกรณ์มหาวิทยาลัย, สถาบันวิทยบริการ, 2536) หมายถึง แฟ้มข้อมูลที่ประกอบขึ้นด้วยกลุ่มของสารนิเทศ (Unit of information) ที่ได้รับการจัดให้อยู่รวมกันตามความต้องการของผู้ใช้แต่ละราย สารนิเทศแต่ละหน่วยในฐานข้อมูล จะประกอบด้วยหน่วยข้อมูล (Data elements) ต่างๆ โดยแต่ละหน่วยข้อมูลแสดงถึง คุณสมบัติเฉพาะของข้อมูล และจัดเก็บไว้ในเขตข้อมูล (Field) โดยเขตข้อมูลทุกค่าจะมีหมายเลขประจำเขตข้อมูล (Field tag) เพื่อใช้ในการอ้างอิงถึง กลุ่มของเขตข้อมูลประกอบกันเป็นระเบียน (Record) ดูรูปที่ 2.1 ประกอบ



รูปที่ 2.1 แสดงฐานข้อมูลซีดีเอส/ไอซีส

โปรแกรมซีดีเอส/ไอซีส มีลักษณะพิเศษเฉพาะแตกต่างจากโปรแกรมคอมพิวเตอร์อื่นๆ กล่าวคือได้รับการออกแบบให้สามารถจัดเก็บค่าข้อมูลและระเบียนข้อมูลที่มีความยาวที่แปรได้ (Variable length) บางเขตข้อมูลอาจไม่มีข้อมูลใดๆ แต่บางเขตข้อมูลอาจมีหน่วยข้อมูลมากกว่าหนึ่งหน่วยก็ได้ เรียกว่าเขตข้อมูลย่อย (Subfield) นอกจากนั้นในบางระเบียนอาจเกิดเขตข้อมูลเดียวกันหลายครั้งได้ เรียกว่า เขตข้อมูลซ้ำ (Repeatable field)

ฐานข้อมูลที่พัฒนาโดยโปรแกรมซีดีเอส/ไอซีส จะประกอบด้วยแฟ้มข้อมูลหลายแฟ้มข้อมูลที่มีความสัมพันธ์กันทางตรรก (Logical relationship) ซึ่งโปรแกรมจะเป็นผู้จัดการระบบแฟ้มข้อมูลนี้ทางกายภาพเอง

โดยผู้ใช้ไม่จำเป็นต้องทราบถึงรายละเอียดของโครงสร้างอย่างละเอียด เป็นระบบที่มีการจัดเก็บข้อมูลได้อย่างมีประสิทธิภาพ และสามารถเข้าถึงข้อมูลได้รวดเร็ว รวมทั้งมีความสามารถในการค้นหาข้อมูลด้วยการใช้ดัชนีพจน์สืบค้น (Search term) โดยใช้หลักการแฟ้มข้อมูลผกผัน (Inverted File)

2.3 ระบบแฟ้มข้อมูลของฐานข้อมูลซีดีเอส/ไอซีส

ระบบแฟ้มข้อมูลของฐานข้อมูลซีดีเอส/ไอซีส (UNESCO, Division of Software Development and Applications, 1989) ประกอบด้วยแฟ้มข้อมูลที่สำคัญ 8 ชนิด (file extension) ดังต่อไปนี้

1. แฟ้มข้อมูลชนิด MST

เป็นแฟ้มข้อมูลหลัก (Master file) สามารถเก็บระเบียนข้อมูลที่มีความยาวที่แปรได้

2. แฟ้มข้อมูลชนิด XRF

เป็นแฟ้มข้อมูลอ้างอิงไข้ว (Cross reference file) ที่เป็นดัชนีชี้ระเบียนของแฟ้มข้อมูลหลักด้วยหมายเลขระเบียน

3. แฟ้มข้อมูลชนิด CNT

เป็นแฟ้มข้อมูลควบคุมบี*ทรี (B*tree) หรือแฟ้มควบคุมพจนานุกรมพจน์สืบค้น (Search term dictionary)

4. แฟ้มข้อมูลชนิด N01

เป็นแฟ้มข้อมูลบี*ทรีโหนด (B*tree node) สำหรับพจน์ที่มีความยาวไม่เกิน 10 ตัวอักระ

5. แฟ้มข้อมูลชนิด L01

เป็นแฟ้มข้อมูลบี*ทรีลีฟ (B*tree leaf) สำหรับพจน์ที่มีความยาวไม่เกิน 10 ตัวอักขระ

6. แฟ้มข้อมูลชนิด N02

เป็นแฟ้มข้อมูลบี*ทรีโหนด (B*tree node) สำหรับพจน์ที่มีความยาวเกิน 10 ตัวอักระ

7. แฟ้มข้อมูลชนิด L02

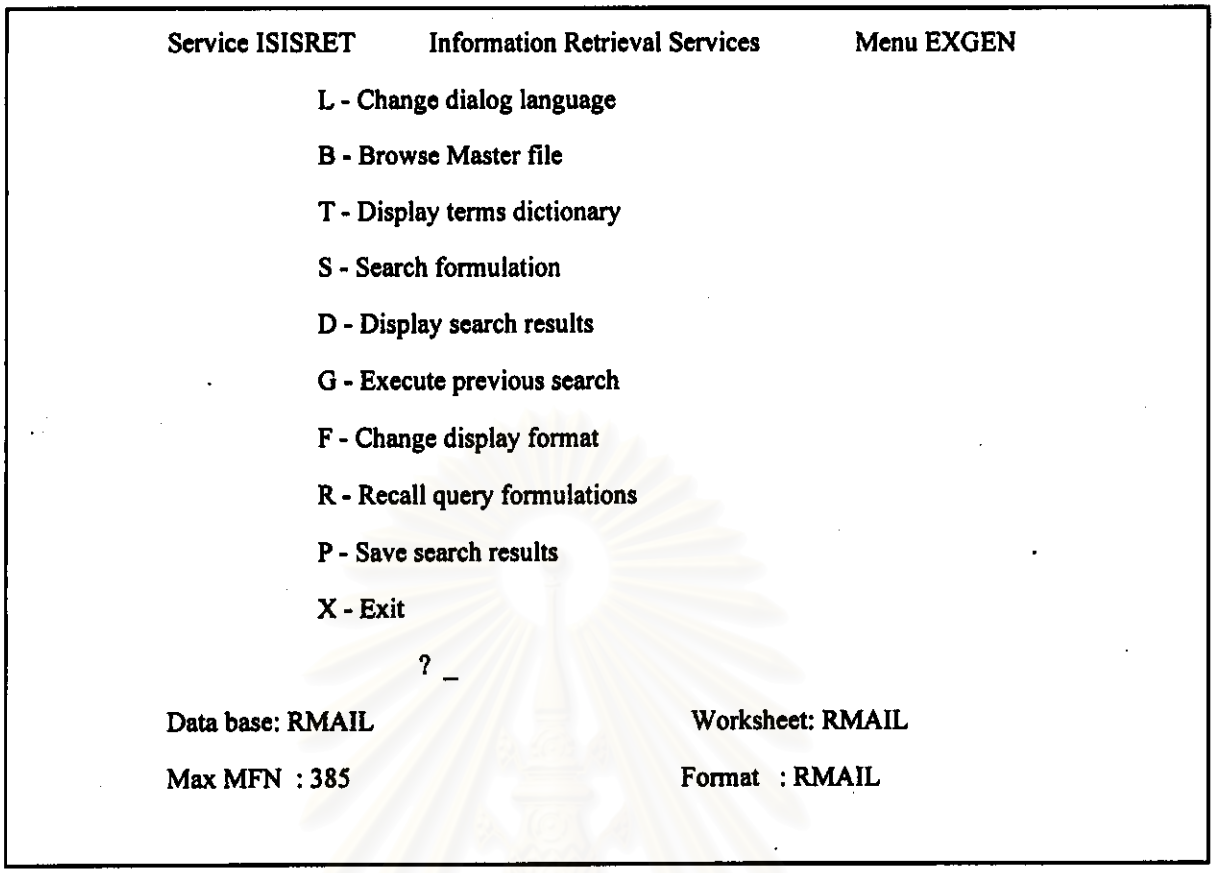
เป็นแฟ้มข้อมูลบี*ทรีลีฟ (B*tree leaf) สำหรับพจน์ที่มีความยาวเกิน 10 ตัวอักขระ

8. แฟ้มข้อมูลชนิด IFP

เป็นแฟ้มข้อมูลโพสติ้งของแฟ้มข้อมูลผกผัน (Inverted file postings)

2.4 โปรแกรมสืบค้นและแสดงผลข้อมูล

(UNESCO, 1989) โปรแกรมซีดีเอส/ไอซีส มีโปรแกรมย่อยที่มีชื่อว่า โปรแกรม ISISRET สำหรับใช้ในการสืบค้นข้อมูล โดยมีลักษณะเมนูดังรูปที่ 2.2



รูปที่ 2.2 แสดงเมนูโปรแกรมย่อย ISISRET

ประกอบด้วยทางเลือกที่สำคัญดังต่อไปนี้

B- Browse Master file

ทางเลือกนี้เป็นการแสดงลักษณะข้อมูลตามลำดับเลขประจำระเบียน โดยเมื่อผู้ใช้กดเลือกทางเลือกนี้ หน้าจอจะขึ้นคำถาม **Starting MFN for browsing ?** หรือ หมายถึงให้ผู้ใช้ใส่หมายเลขระเบียนที่ต้องการให้ระบบนำมาแสดง ในกรณีนี้ผู้ใช้สามารถเลือกปฏิบัติดังนี้

1. ใส่เลขระเบียนที่ต้องการให้ระบบเริ่มต้นแสดงให้ดู เช่น 1
2. กดปุ่ม **Enter** จะหมายถึง ให้ระบบแสดงตั้งแต่เลขระเบียนที่ 1

T - Display terms dictionary

เป็นการให้ระบบแสดงพจนานุกรมพจน์สืบค้นทั้งหมดที่ถูกจัดเก็บไว้ในแฟ้มข้อมูลผกผัน ตามข้อกำหนดในตาราง FST เมื่อผู้ใช้เลือกทางเลือกนี้ จะปรากฏข้อความว่า

key : ?

ความหมายคือ ให้ระบบแสดงพจน์สืบค้นใดเป็นคำแรก คำหรืออักษรใดๆที่ป้อนไป ณ จุดนี้คือ จุดเริ่มต้นของการแสดงรายการพจน์สืบค้น เช่น

key : energy

ระบบจะเริ่มแสดงคำค้นบนจอภาพโดยเริ่มที่อักษร ENERGY หากต้องการให้ระบบแสดงคำค้นตั้งแต่ต้นรายการ เมื่อปรากฏคำว่า key : ? ให้กดปุ่ม Enter

การแสดงรายการพจน์สืบค้นของระบบบนจอภาพนั้น จะแสดงเป็น 2 แถว เป็น จำนวน 36 รายการ ข้างหน้าแต่ละรายการจะมีเครื่องหมาย - นำหน้า แสดงตัวอย่างดังรูปที่ 2.3

```
Dictionary list                          Data Base Name: ENGY
==========================================================================
_ ENERGY                                 _ ENERGY AND ENVIRONMENT
_ ENERGY  RESOURCES                      _ ENERGY AND ENVIRONMENTAL BALAN
_ ENERGY & ENVIRONMENTAL IMPLICA         _ ENERGY AND ENVIRONMENTAL ISSUE
_ ENERGY 2000                            _ ENERGY AND ENVIRONMENTAL UPDAT
_ ENERGY 2000-2020                       _ ENERGY AND MATERIALS USE IN TH
_ ENERGY ABSTRACTS FOR POLICY AN         _ ENERGY AND MINERAL DIVISION, U
_ ENERGY ADJUSTMENT CHARGE               _ ENERGY AND MINERALS .
_ ENERGY ALTERNATIVES IN SRI LAN         _ ENERGY AND NUCLEAR POWER PLANN
_ ENERGY ANALYSIS IN A TEXTILE F         _ ENERGY AND OTHER DEVELOPMENT C
_ ENERGY ANALYSIS IN RURAL REGIO         _ ENERGY AND RESOURCE RECOVERY F
_ ENERGY ANALYSIS OF MAIZE PRODU         _ ENERGY AND RURAL DEVELOPMENT
_ ENERGY ANALYSIS OF POTATO PROD         _ ENERGY AND RURAL DEVELOPMENT R
_ ENERGY ANALYSIS OF PROCESSES O         _ ENERGY AND SECURITY
_ ENERGY ANALYSIS OF THE INFLUEN         _ ENERGY AND STRUCTURAL CHANGE I
_ ENERGY AND AGRICULTURE IN MALA         _ ENERGY AND SULFER DIOXIDE ANAL
_ ENERGY AND DEVELOPMENT                 _ ENERGY AND TECHNOLOGY IN ARCHI
_ ENERGY AND ECONOMY GLOBAL INTE         _ ENERGY AND THE DEVELOPING NATI
_ ENERGY AND ELECTRICITY CEMAND          _ ENERGY AND THE ENVIRONMENT
Key  : energy□
```

รูปที่ 2.3 การแสดงพจนานุกรมคำค้นในระบบเดิม

การควบคุมตัวชี้ตำแหน่ง (cursor) และจอภาพทำได้โดย (จุฬาลงกรณ์มหาวิทยาลัย, สถาบันวิทยบริการ, 2536)

ปุ่ม PgDn จะบังคับให้จอภาพเลื่อนไปแสดงหน้าต่อไป ระบบจะแสดงตามลำดับและไม่สามารถย้อนกลับได้

ปุ่มลูกศรทั้งสี่, ปุ่ม Home, ปุ่ม End

ควบคุมตัวชี้ตำแหน่งให้เลื่อนไปยังตำแหน่งของคำค้นที่ผู้ใช้ต้องการเลือกค้นได้

ปุ่ม S เป็นการเลือกรายการคำค้นคำนั้นๆ (ตามปกติเมื่อกด S หน้ารายการใด จะปรากฏแสงสว่างหน้ารายการคำนั้นๆ) และในกรณีที่ผู้ใช้เลือกรายการคำค้นมากกว่า 1 คำ โดยกด S หน้ารายการเหล่านั้น ระบบจะถือว่าผู้ใช้เชื่อมรายการคำค้นนั้นๆด้วยคำเชื่อม OR

คำเชื่อมต่างๆ (Operators)

ได้แก่ + * ^ $ ผู้ใช้สามารถป้อนคำเชื่อมเหล่านี้ข้างหน้ารายการคำค้นแต่ละคำเพื่อสร้าง search expression แต่ละครั้งได้

ปุ่ม C เป็นการออกจากการทำงานในทางเลือกนี้ และยกเลิกการทำงาน ณ ขณะนั้นด้วย

ปุ่ม X เป็นการออกจากการทำงานขณะนี้ พร้อมให้ระบบทำงานตามคำสั่งหรือทางเลือกใดๆ ที่ได้เลือกไว้ตามคำค้นต่างๆ (หากไม่มีรายการคำค้นใดๆ เลือกไว้ระบบจะกลับคืนสู่รายการเมนูของ ISISRET)

S - Search formulation

เป็นการนำผู้ใช้เข้าสู่ขั้นตอนแรกของการสืบค้น นั่นคือ เมื่อเลือก S หน้าจอภาพจะปรากฏคำว่า **Search expression ?**

สิ่งที่ผู้ใช้ต้องปฏิบัติคือ ใช้ภาษาการสืบค้นตามเงื่อนไขที่ต้องการ นิพจน์หรือข้อมูลการสืบค้นแต่ละคำถามจะยาวไม่เกิน 250 อักขระ เมื่อผู้ใช้กำหนดนิพจน์ตามต้องการแล้ว เช่น ENERGY+THAILAND และกดปุ่ม **Enter** เพื่อให้ระบบทำงาน คำตอบที่ระบบแสดงบนจอภาพจะเป็นรูปดังต่อไปนี้

```
Set 2:   ENERGY + THAILAND
P=   158   ENERGY
P=   140   THAILAND
T=   280  -  #3: ENERGY + THAILAND
T=   280  -  #2: #3
```

รูปที่ 2.4 แสดงผลลัพธ์การสืบค้น ENERGY + THAILAND

Set n หมายถึงหมายเลขนิพจน์คำถาม (Search expression) ซึ่งระบบเป็นผู้กำหนดให้แต่ละครั้ง และผู้ใช้สามารถเรียกหมายเลขดังกล่าวใช้งานได้ต่อไป

P หมายถึง จำนวนครั้งที่พจน์สืบค้นนั้นปรากฏในฐานข้อมูล

T หมายถึง จำนวนระเบียนทั้งหมดที่มีคำของนิพจน์คำถามนั้น

การสืบค้นแบบรวมหลายนิพจน์และมีการเชื่อมนิพจน์ ระบบจะแสดงจำนวนของแต่ละนิพจน์เรียงมาก่อนแล้วจึงแสดงผลลัพธ์ของจำนวนระเบียนที่มีข้อมูลตามเงื่อนไข

เมื่อระบบแสดงผลลัพธ์แล้ว ผู้ใช้สามารถเรียกดูข้อมูลได้เลย โดยใช้ทางเลือก D หรือ หากกด ปุ่ม Enter ระบบจะคืนสู่จอภาพเมนู EXGEN ตามเดิมแล้วจึงเลือก D ก็ได้

D - Display search results

เมื่อผู้ใช้เลือกสืบค้นตามวิธีต่างๆ และได้ทราบว่าการสืบค้นของตนพบในที่ระเบียนแล้วหากผลลัพธ์ของการสืบค้นครั้งนั้น ผู้ใช้สามารถกด D ได้เลย ระบบจะแสดงผลลัพธ์การสืบค้นครั้งล่าสุด แต่หากการสืบค้นครั้งล่าสุดนั้นไม่มีผลลัพธ์ ระบบจะแสดงบนจอภาพดังนี้

***** End of display *****

โดยปกติระบบจะแสดงผลลัพธ์การสืบค้นตามลำดับระเบียนของข้อมูลผลลัพธ์นั้น หากมากกว่า 1 จอภาพ ระบบจะแสดงให้ทราบด้วย ...more ผู้ใช้สามารถกดปุ่ม Enter เพื่อให้จอภาพเลื่อนไปยังส่วนต่อไปหรือหากผู้ใช้ต้องการ หยุดเพียงเท่านั้นและต้องการออกจากการแสดงผลสามารถกด X ออกมาสู่เมนู ISISRET ได้

G - Execute previous search

ในหลายกรณีผู้ใช้ต้องการสืบค้นแต่ละครั้งเป็นพื้นฐานการสืบค้นครั้งต่อๆไป ทางเลือกนี้เป็นการเรียกการสืบค้นครั้งก่อนๆมาดำเนินการใหม่

F - Change display format

ในการให้ระบบแสดงรูปแบบการสืบค้นนั้น บางกรณีผู้ใช้อาจต้องการเปลี่ยนแปลงรูปแบบการแสดงผลเป็นแบบอื่นๆ ตามที่กำหนดไว้แล้ว โดยอ้างอิงถึงรูปแบบนั้นด้วย @ นำหน้าชื่อรูปแบบดังกล่าว หรือ เป็นรูปแบบใหม่ที่เปลี่ยนแปลงขณะนั้นก็ได้

R - Recall query formulation

ทางเลือกนี้เป็นการเรียกแสดงแบบการสืบค้นทั้งหมดโดยแสดงประวัติการสืบค้นของผู้ใช้ทั้งหมด โดยจะแสดงหมายเลขเซตตั้งแต่เซตแรกของการเรียกใช้โปรแกรม ISISRET ขณะนั้นจนถึงหมายเลขเซตล่าสุด พร้อมทั้งแสดงลักษณะของรายการคำถามหรือรายการนิพจน์การสืบค้น (search expression) แต่ละรายการและจำนวนผลลัพธ์การสืบค้นแต่ละครั้งนั้นด้วย

P - Save search results

เมื่อเสร็จสิ้นการสืบค้นแต่ละครั้ง ผู้ใช้สามารถจะเลือกได้ว่าจะเก็บผลลัพธ์การสืบค้นนั้นหรือไม่ หากต้องการเก็บผลลัพธ์ก็สามารถเก็บบันทึกลงแฟ้มข้อมูล เพื่อนำกลับมาใช้ใหม่ได้

จะเก็บผลลัพธ์การสืบค้นด้วยทางเลือก Save search results และนำผลลัพธ์ไปพิมพ์ออกทางเครื่องพิมพ์ด้วยเมนู Print services

X - Exit

ทางเลือกนี้ใช้ในการออกจากโปรแกรม ISISRET คืนสู่โปรแกรมหลักของซีดีเอส/ไอซีส

2.5 <u>การสืบค้นข้อมูล</u>

(UNESCO, 1989) ภาษาการสืบค้นของโปรแกรมซีดีเอส/ไอซีสใช้พื้นฐานของพีชคณิตบูลีน (boolean algebra) เพื่อช่วยให้สามารถแสดงพจน์สืบค้นที่ระบบได้สร้างเป็นดัชนีไว้ให้ โปรแกรมซีดีเอส/ไอซีสได้กำหนดทางเลือก T - Display terms dictionary ในเมนู Information Retrieval Services ซึ่งเป็นการแสดงรายการพจน์ที่ผู้ใช้สามารถเลือกสืบค้นจากรายการดังกล่าว มาประกอบสร้างเป็นรูปแบบการสืบค้นให้ตรงตามความต้องการของผู้ใช้ได้

สำหรับกฎเกณฑ์และเทคนิคการสืบค้นที่ควรทราบมีดังนี้

1. <u>ประเภทของพจน์สืบค้น</u>

1.1 <u>พจน์เฉพาะเจาะจง</u> (Precise term)

ได้แก่ พจน์หรือข้อความใด ๆ ที่สามารถสืบค้นได้และได้กำหนดไว้ในฐานข้อมูลแล้ว พจน์เหล่านี้จะแสดงให้เห็นในทางเลือก T - Display terms dictionary ของโปรแกรม ISISRET การสืบค้นด้วยพจน์ที่เฉพาะเจาะจงนี้มีข้อกำหนดเพียงว่า ผู้ค้นต้องระบุพจน์หรือข้อความตามที่กำหนดไว้เพื่อการสืบค้นเท่านั้น หากใช้พจน์หรือข้อความที่สะกดไม่ถูกต้อง ระบบการสืบค้นจะไม่สามารถค้นให้ได้ เช่น กรณีต้องการสืบค้นพจน์ COLOURS แต่ในฐานข้อมูลเก็บภายใต้พจน์ COLORS เมื่อเราสืบค้นพจน์ COLOURS หน้าจอจะขึ้น NOT FOUND ดังนั้นการสืบค้นด้วยพจน์ที่เฉพาะเจาะจงจะต้องระมัดระวังการใช้เป็นพิเศษ และข้อพึงระวังอีกประการคือ หากพจน์ที่สืบค้นมีเครื่องหมายวงเล็บ ( ) หรือเครื่องหมายเพื่อการสืบค้นใด ๆ เช่น * + $ หรือ ^ เป็นส่วนประกอบ ต้องใช้เครื่องหมายคำพูด (") คร่อมคำสืบค้นนั้น เช่น คำสืบค้นเป็น "GERMANY (FEDERAL REPUBLIC)" มิเช่นนั้นจะปรากฏข้อความว่ารูปแบบที่สืบค้นผิดพลาด

1.2 <u>การค้นแบบตัดปลายขวา</u> (Right-Truncated)

ลักษณะการค้นในข้อนี้จะต่างจากการค้นแบบพจน์เฉพาะเจาะจงในข้างต้น นั่นคือ แทนที่จะต้องระบุพจน์สืบค้นให้ถูกต้อง ผู้ใช้สามารถป้อนเพียงส่วนของพจน์ที่จะค้นหรือรากของพจน์นั้น แล้วใช้เครื่องหมาย $ ต่อท้ายทันที ระบบจะสืบค้นพจน์ที่เริ่มต้นเช่นเดียวกับพจน์ที่ผู้ใช้กำหนด ทั้งหมดเท่าที่ปรากฏในแฟ้มข้อมูล เช่น พจน์ BUILDINGS$ ระบบจะค้นที่พจน์ต่อไปนี้ให้

BUILDINGS
BUILDINGS AND PROJECTS
BUILDINGS AND THE ARTS

นั่นคือ ระบบจะค้นและนำพจน์ใด ๆ ก็ตามที่ขึ้นต้นตามเงื่อนไข คือพจน์ BUILDINGS และไม่ว่าจะมีเครื่องหมายหรืออักษรใด ๆ ต่อท้าย BUILDINGS ก็จะถูกค้นขึ้นมาให้ด้วยเช่นกัน

2. <u>การใช้คำเชื่อมหรือตัวดำเนินการสืบค้น</u> (Search Operators)

ในการสืบค้นข้อมูลแต่ละครั้ง ผู้ใช้อาจรวมพจน์ตั้งแต่ 2 พจน์ขึ้นไปได้โดยใช้เครื่องหมายหรือสัญลักษณ์เชื่อมความสัมพันธ์ระหว่างพจน์ เครื่องหมายเหล่านั้นจะแทนความหมาย ดังนี้

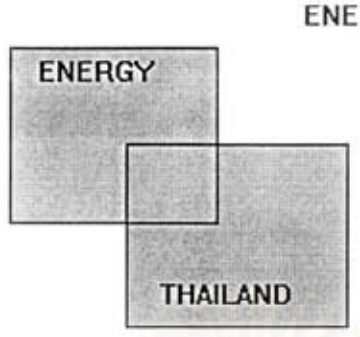
2.1 "หรือ" (OR) แทนด้วยเครื่องหมาย + การสืบค้นด้วยคำเชื่อม "หรือ" นี้ จะขยายขอบเขตของการสืบค้นครั้งนั้น ๆ

2.2 "และ" (AND) แทนด้วยเครื่องหมาย * การสืบค้นด้วยคำเชื่อม "และ" นี้ จะจำกัดขอบเขตของการสืบค้นครั้งนั้น ๆ

2.3 "ไม่" (NOT) แทนด้วยเครื่องหมาย ^ การสืบค้นด้วยคำเชื่อม "ไม่" นี้ จะต้องกระทำด้วยความระมัดระวังเป็นพิเศษ เนื่องจากอาจเป็นผลทำให้ข้อมูลในส่วนที่เกี่ยวข้องสูญเสียไป

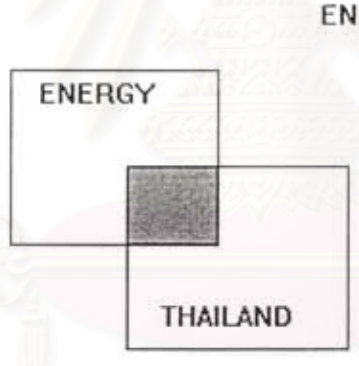
ดูรูปประกอบการแสดงการทำงานตามพีชคณิตบูลลีน "หรือ" "และ" "ไม่" ดังต่อไปนี้

จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 2.5 ตรรก "หรือ"

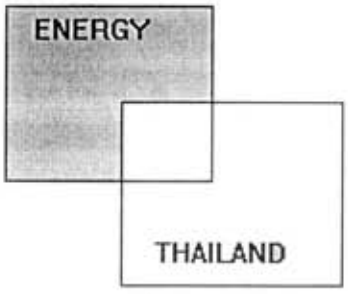
ตามรูปที่ 2.5 ส่วนของ ENERGY และส่วนของ THAILAND เมื่อค้นตามตรรก "หรือ" (+) แล้วส่วนของ ENERGY และส่วนของ THAILAND รวมทั้งส่วนที่มี ENERGY และ THAILAND ร่วมกัน คือบริเวณพื้นที่สีเข้มทั้งหมด จะถูกระบบดึงขึ้นมาเป็นคำตอบแก่ผู้ใช้



รูปที่ 2.6 ตรรก "และ"

ตามรูปที่ 2.6 ตรรก "และ" (*) ต้องการคำตอบที่มีทั้ง ENERGY และ THAILAND ระบบจะดึงเฉพาะส่วนที่มีคำตอบร่วมกันของทั้ง ENERGY และ THAILAND คือ ส่วนที่เป็นตารางตรงกลางหรือบริเวณพื้นที่สีเข้มเท่านั้น

ENERGY^THAILAND



รูปที่ 2.7 ตรรก "ไม่"

ตามรูปที่ 2.7 ตรรก "ไม่" (^) ต้องการเฉพาะส่วนของ ENERGY ไม่รวม THAILAND และไม่รวมส่วนที่ร่วมกันของทั้ง ENERGY และ THAILAND คือบริเวณพื้นที่สีเข้มเท่านั้น การใช้คำเชื่อม ^ นี้ไม่เหมือนการใช้คำเชื่อม * หรือ + เพราะคำเชื่อม * หรือ + จะให้ผลการสืบค้นเหมือนกันในกรณีที่จะสลับที่กัน แต่สำหรับคำเชื่อม ^ หากนิพจน์ระหว่าง ^ สลับที่กัน ผลลัพธ์ที่ได้จะต่างกัน เช่น ENERGY^THAILAND ผลลัพธ์จะแตกต่างจาก THAILAND^ENERGY

3. <u>ระดับการสืบค้นของคำเชื่อม</u> (Operator priority)

เนื่องจากในการสืบค้นแต่ละครั้ง ผู้ใช้สามารถประกอบพจน์ค้นเป็นข้อความที่มีเครื่องหมายต่าง ๆ มากกว่า 1 อย่างได้ การจัดลำดับการดำเนินการของระบบถือตามกฎพีชคณิตปกติ คือ สามารถใช้เครื่องหมายวงเล็บ เพื่อบังคับให้จัดการตามคำสั่งในวงเล็บนั้นก่อน แต่ในกรณีปกติ ระบบจะจัดการคำสั่งตามลำดับ ดังนี้

สูงสุด S

* ^

ต่ำสุด +

ในกรณีที่คำเชื่อม 2 คำ ซึ่งมีระดับการทำงานอยู่ระดับเดียวกัน อยู่ในวงเล็บ หรือขั้นตอนการทำงานเท่ากันระบบจะทำงานจากซ้ายไปขวา

เช่น A + B + C

ระบบจะทำงานตรง A + B ก่อน จากนั้นจึงจะทำงานตามคำเชื่อม + ระหว่างผลลัพธ์ของ A + B กับ C

ในกรณีที่คำเชื่อม 2 คำ ซึ่งมีระดับการทำงานอยู่แตกต่างกันระบบจะทำงานที่ระดับการสืบค้นสูงกว่าก่อน

เช่น A + B * C

ระบบจะทำงานตรง B * C ก่อน จากนั้นจึงจะทำงานตามคำเชื่อม + ระหว่าง A กับผลลัพธ์ของ B * C แต่หากคำถามเป็นเช่นนี้ (A + B) * C ระบบจะทำงาน (A + B) ก่อนจากนั้นจึงจะนำผลลัพธ์ที่ได้มาจัดการตามคำเชื่อม * กับ C

ข้อควรระวัง

1. คำเชื่อมหรือตัวดำเนินการสืบค้นจะอยู่ติดกันไม่ได้
2. จำนวนวงเล็บที่ใช้ในนิพจน์การสืบค้นแต่ละครั้งต้องมีจำนวนวงเล็บเปิดและวงเล็บปิดเป็นคู่ๆ

4. การกำหนดการสืบค้นตามเขตข้อมูลหรือกลุ่มเขตข้อมูล (Operand qualifier)

ในบางกรณี การสืบค้นอาจจำเป็นต้องระบุเขตข้อมูลเพื่อให้ได้ข้อมูลที่ตรงตามความต้องการที่สุดการกำหนดเขตข้อมูลมีรูปแบบดังนี้

พจน์สืบค้น/(t1, t2, t3,....)

นั่นคือ ต้องการสืบค้นพจน์ใด กำหนดให้อยู่ในเขตข้อมูลใด โดย t1, t2, t3 นั้น คือ เลขหรือกลุ่มของ field identifiers ที่ถูกกำหนดรูปแบบการจัดทำดัชนีในตารางเลือกเขตข้อมูล (File Select Table - FST) ไว้แล้ว

5. การแสดงประวัติและวิธีการสืบค้น

แต่ละครั้งที่ผู้ใช้สืบค้นด้วยพจน์ใด ๆ คำถามหรือนิพจน์การสืบค้นนั้น ๆ จะถูกนับเป็นชุด ๆ และบันทึกประวัติการสืบค้นไว้เพื่อการเรียกแสดงใหม่ การอาศัยแบบการสืบค้นเดิมหรือเพื่อนำมาปรับแบบการสืบค้นใหม่

ตัวอย่าง กำหนดสูตรการสืบค้นเป็น

(WATER+ENERGY)*THAILAND

CDS/ISIS จะแสดงขั้นตอนการทำงาน ดังนี้

บรรทัดที่

1 Set 1: (WATER+ENERGY)*THAILAND

2 P= 1 WATER

3 P= 158 ENERGY

4 T= 157 - #2: WATER + ENERGY

5 P= 140 THAILAND

6 T= 10 - #3: #2 * THAILAND

7 T= 10 - #1: #3

<u>คำอธิบาย</u>

1 Set n หมายถึงหมายเลขนิพจน์การสืบค้นหรือสูตรการสืบค้นที่ระบบกำหนดให้แต่ละครั้งที่ทำการสืบค้น หมายเลขดังกล่าวสามารถใช้อ้างถึง เพื่อนำการสืบค้นเดิมมาปรับปรุงแก้ไขใหม่ได้

2-3 P หมายถึงจำนวนครั้งที่พจน์สืบค้นนั้น ๆ ปรากฏในฐานข้อมูล

4 T หมายถึงจำนวนระเบียนที่มีข้อมูลภายใต้เงื่อนไขของสูตรการสืบค้นย่อยแต่ละชุด ในกรณีนี้สูตรการสืบค้นย่อย WATER + ENERGY ถือเป็นสูตรการสืบค้นย่อยที่ 2

5 P หมายถึงจำนวนครั้งที่พจน์สืบค้นถัดมาปรากฏในฐานข้อมูล

6 เป็นจำนวนระเบียนของสูตรการสืบค้นย่อยที่ 3 คือ ประมวลผลรวมของสูตรการสืบค้นย่อยที่ 2 (บรรทัดที่ 4) กับพจน์สืบค้น THAILAND (บรรทัดที่ 5)

7 เป็นจำนวนระเบียนทั้งหมดที่พบตามเงื่อนไขของสูตรการสืบค้น ดังนี้ (หมายเลขนิพจน์คำถามที่ 1) ซึ่งมีผลลัพธ์เท่ากับจำนวนระเบียนของสูตรการสืบค้นย่อยที่ 3 (บรรทัดที่ 6)

เมื่อระบบประมวลผลจะได้ผลลัพธ์ที่แท้จริง ของตัวอย่างสูตรการสืบค้นนี้แล้ว โปรแกรมถือว่าการสืบค้นที่แท้จริงมีเพียงชุดเดียว คือ

Set 1 : (WATER+ENERGY)*THAILAND

ดังนั้นเมื่อผู้ใช้ต้องการสืบค้นเรื่องดังกล่าวให้เฉพาะเจาะจงอีก สามารถทำได้โดยอ้างถึง Set 1 นี้ด้วยเครื่องหมาย # ได้หลายแบบ เช่น

#1^GAS หรือ #1/(64)

อย่างไรก็ตาม สำหรับผู้ที่ยังไม่มีประสบการณ์หรือยังไม่คุ้นเคยกับเทคนิควิธีสืบค้นต่าง ๆ สามารถสืบค้นได้อย่างง่าย ๆ ทีละพจน์หรือใช้คำเชื่อม + * ^ เท่านั้น ก็สามารถได้ผลลัพธ์ที่ตรงกัน เช่น ยกตัวอย่างนิพจน์การสืบค้นเป็น (WATER+ENERGY)*THAILAND^GAS ผู้ใช้สามารถค้นทีละครั้ง รวม 4 ครั้ง ดังนี้

Set 1: WATER + ENERGY

Set 2: THAILAND

Set 3: GAS

Set 4: #1 * #2 ^ #3

2.6 การกำหนดรูปแบบข้อมูลเพื่อการแสดงผล

(UNESCO, 1989) ภาษารูปแบบการแสดงผล(The Display Formatting Language) ของโปรแกรมซีดีเอส/ไอซีส ประกอบด้วย ชุดของคำสั่ง (commands) หรือข้อความสั่ง (statements) ตั้งแต่หนึ่งชุดขึ้นไปโดยพารามิเตอร์ต่างๆ แต่ละชุดจะถูกคั่นด้วยเครื่องหมายจุลภาค (,) หนึ่งหรือหลายๆตัว หรือเว้นว่างเพื่อช่วยให้อ่านและเข้าใจง่ายขึ้น ใช้สำหรับกำหนดรูปแบบของข้อมูลในแต่ละระเบียนในฐานข้อมูลตามความต้องการของผู้ใช้

สำหรับภาษาการกำหนดรูปแบบการแสดงผลที่ควรทราบมีดังนี้

1. พารามิเตอร์เลือกเขตข้อมูล (Field Selectors)

เป็นคำสั่งที่ใช้ในการระบุค่าของข้อมูลในแต่ละเขตข้อมูล (field) หรือเขตข้อมูลย่อย (subfield) ของระเบียน (record) หนึ่งๆ รวมไปถึงคำสั่งระบุค่า MFN ด้วย

1.1 คำสั่งเขตข้อมูล (Field command)

รูปแบบ:	Vttt
V	คือ รหัสของเขตข้อมูลเพื่อบอกว่าเป็นพารามิเตอร์ของเขตข้อมูลที่มีความยาวแปรผัน(variable length record)
ttt	คือ ตัวเลขประจำเขตข้อมูล (Field tag) ที่ต้องการ
ตัวอย่าง	V100 หรือ v100

1.2 คำสั่งเขตข้อมูลย่อย (Subfield command)

รูปแบบ:	^x
^x	คือรหัสของเขตข้อมูลย่อย(subfield code) ที่ต้องการ
ตัวอย่าง	V100^a หรือ v100^a

1.3 คำสั่งย่อหน้า (Indentation command)

รูปแบบ:	(f,c) หรือ (f)
f	แสดงจำนวนช่องว่างที่ต้องการให้เว้นทางซ้ายมือ ก่อนที่จะแสดงค่าของข้อมูลบรรทัดแรกของเขตข้อมูลที่ต้องการ (first line of the field)
c	แสดงจำนวนช่องว่างที่ต้องการให้เว้นทางซ้ายมือ ก่อนที่จะแสดงค่าของข้อมูลบรรทัดต่อๆไปของเขตข้อมูลที่ต้องการ(all continuation lines) ใช้ในกรณีที่ค่าของเขตข้อมูลมีความยาวเกินหนึ่งบรรทัด
ตัวอย่าง	V100 หรือ v100
	V100(10) หรือ v100(10)
	V100(0,8) หรือ v100(0,8)

1.4 คำสั่งหมายเลขระเบียน (MFN command)

รูปแบบ:	MFN หรือ MFN(d)
d	ระบุจำนวนหลักของ MFN ถ้าไม่กำหนด โปรแกรมจะแสดงผลลัพธ์เป็นเลข 6 หลัก
ตัวอย่าง	MFN หรือ mfn
	MFN(4) หรือ mfn(4)

2. คำสั่งกำหนดแบบ (Mode command)

การกำหนดรูปแบบของโปรแกรมสามารถกระทำได้ 3 แบบ (mode) ดังนี้

2.1 แบบพิสูจน์อักษร (Proof mode)

การทำงานของแบบวิธีนี้ โปรแกรมจะแสดงค่าของข้อมูล ในลักษณะที่เหมือนกับที่มีอยู่ในเขตข้อมูลนั้นจริงๆ ใช้สำหรับตรวจสอบแก้ไขข้อมูล (proof-reading)

2.2 แบบหัวข้อ (Heading mode)

การทำงานของแบบวิธีนี้ โปรแกรมจะแสดงค่าของข้อมูล โดยจะแทนที่รหัสแสดงเขตข้อมูลย่อยด้วยเครื่องหมายวรรคตอน รวมทั้งอักขระควบคุมต่างๆ (control characters) ที่แฝงอยู่ในข้อมูล (เช่น % < และ >) จะถูกตัดทิ้งไป

2.3 <u>แบบข้อมูล</u> (Data mode)

การทำงานของแบบวิธีนี้ โปรแกรมจะแสดงค่าของข้อมูลในลักษณะเดียวกับแบบหัวข้อ จะต่างกันเพียงแต่ โปรแกรมจะเติมเครื่องหมายมหัพภาค (.) ตามด้วยช่องว่างอีก 2 ตำแหน่ง เมื่อจบค่าข้อมูลแต่ละเขตข้อมูล (field) หรือเพียงเติมช่องว่าง 2 ตำแหน่งให้ ถ้ามีเครื่องหมายมหัพภาคอยู่แล้วเมื่อจบขตข้อมูลนั้นๆ

รูปแบบ:	**Mmc**	
m	ใช้กำหนดแบบที่ต้องการได้ ดังนี้	
	P	proof mode
	H	heading mode
	D	data mode
c	ใช้กำหนดการเปลี่ยนเป็นตัวอักษรเป็นตัวใหญ่ ตัวเล็ก ดังนี้	
	U	ข้อมูลเปลี่ยนเป็นตัวอักษรตัวใหญ่ทั้งหมด
	L	ข้อมูลไม่เปลี่ยนแปลง

การใช้คำสั่งเปลี่ยนแปลงแบบวิธีดังกล่าวข้างต้น สามารถใช้ได้บ่อยตามความจำเป็น ถ้าหากผู้ใช้ลืมกำหนด โปรแกรมจะจักการรูปแบบของข้อมูลด้วยรูปแบบ 'MPL'

3. <u>คำสั่งเว้นระยะแนวนอนและแนวตั้ง</u> (Horizontal and vertical spacing commands)

Xn	แทรกช่องว่างจำนวน n ก่อนการจัดรูปแบบเขตข้อมูลถัดไป
Cn	กำหนดจุดวางตัวอักษร ณ ตำแหน่ง n
/	ขึ้นบรรทัดใหม่
#	ขึ้นบรรทัดใหม่(แบบไม่มีเงื่อนไข)

4. <u>สัญพจน์</u> (Literals)

4.1 <u>สัญพจน์แบบมีเงื่อนไข</u> (Conditional literals)

รูปแบบ "....."

4.2 <u>สัญพจน์ต่อท้ายเขตข้อมูลซ้ำ</u> (Suffix repeatable literals)

รูปแบบ +|.....|

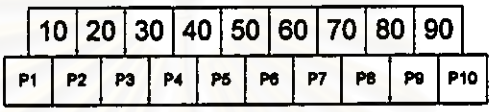
4.3 <u>สัญพจน์แบบไม่มีเงื่อนไข</u> (Unconditional literals)

รูปแบบ '.....'

2.7 โครงสร้างข้อมูลแบบบี*ทรีและการค้นหาข้อมูล

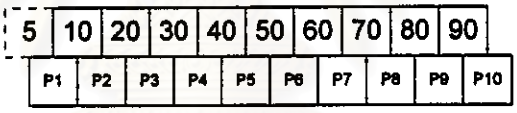
โครงสร้างข้อมูลแบบบี*ทรี เป็นโครงสร้างข้อมูลที่ปรับจากโครงสร้างข้อมูลแบบบีทรี ให้มีคุณสมบัติเพิ่มเติมที่เรียกว่า เทคนิคโอเวอร์โฟลว์ (Overflow Technique) เพื่อพัฒนาขั้นตอนวิธีการเพิ่มข้อมูล โดยนำข้อมูลที่เต็มจากโหนดหนึ่ง ไปยังโหนดพี่น้องซึ่งยังไม่เต็ม แทนที่จะสร้างโหนดใหม่ ความแตกต่างจากบีทรีก็คือ ทุกๆ โหนดจะมีข้อมูลอยู่อย่างน้อย 2 ใน 3 แทนที่จะเป็นครึ่งหนึ่ง ทำให้ประหยัดเนื้อที่กว่าบีทรีและลดจำนวนโหนดลง ทำให้การค้นหาเร็วขึ้น (อิศระ เดชะไกศยะ, 2530)

ปกติโครงสร้างข้อมูลแบบบีทรีและแบบบี*ทรี ชนิด m ทาง จะมีจำนวนคำหลักคือ m-1 คำ

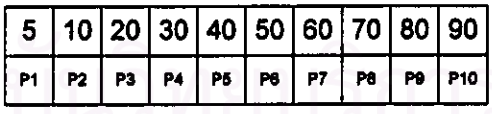


รูปที่ 2.8 แสดงโหนดของบี*ทรีชนิด 10 ทาง โดยมีคำหลัก 9 คำ

โหนดของบี*ทรี ที่ใช้ในฐานข้อมูลซีดีเอส/ไอซีสจะเป็นชนิด 10 ทางโดยมีการเพิ่มคำหลักอีก 1 คำ



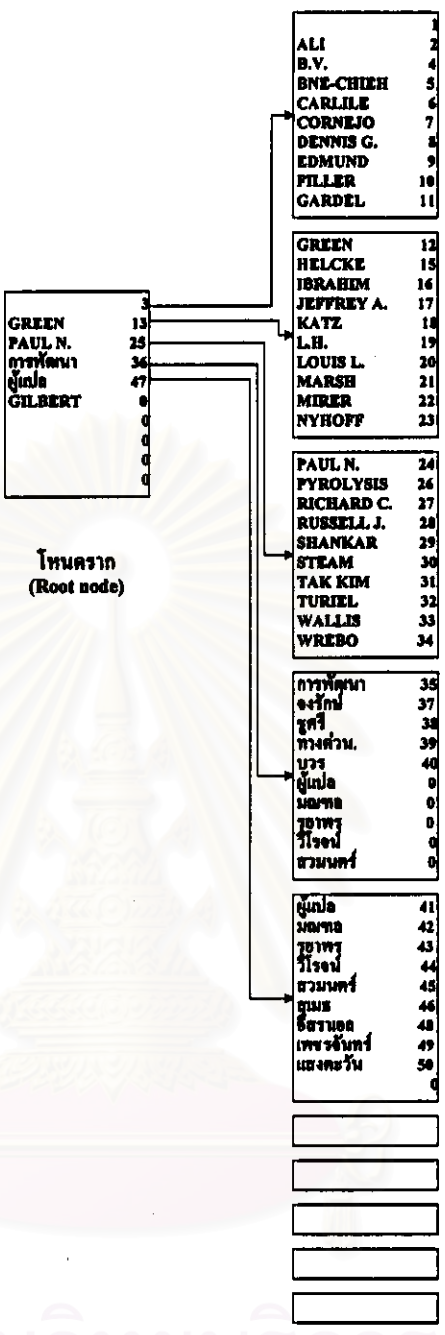
รูปที่ 2.9 แสดงโหนดของบี*ทรีชนิด 10 ทางที่มีการเพิ่มคำหลักอีก 1 คำ



รูปที่ 2.10 แสดงการจับคู่ของคำหลักและตัวชี้ภายใน 1 โหนดของบี*ทรี

โครงสร้างบี*ทรีที่ใช้เป็นพื้นฐานของฐานข้อมูลซีดีเอส/ไอซีส จะมีจำนวนคำหลัก (key) สูงสุดในแต่ละโหนดจำนวน 10 คำ (UNESCO, 1989) และมีเส้นทางมากที่สุด 10 เส้นทาง แสดงตัวอย่างดังรูปต่อไปนี้

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 2.11 แสดงตัวอย่างบี*ทรีของฐานข้อมูลซีดีเอส/ไอซีส

สำหรับการค้นหาข้อมูลแบบบี*ทรีของฐานข้อมูลซีดีเอส/ไอซีส (UNESCO, 1989) เมื่อ ITEM เป็นข้อมูลที่ต้องการ อัลกอริทึมต่อไปนี้จะทำการค้นหาตำแหน่งของ ITEM ที่อยู่ในบี*ทรี

1. เปรียบเทียบ ITEM กับคำหลักทั้ง 10 คำในโหนด N ของบี*ทรี โดยเริ่มจากโหนด N ที่เป็นโหนดรากก่อน

1.1 ถ้า ITEM > หรือ = คำหลักตัวที่ 1 แต่ < คำหลักตัวที่ 2 ของโหนด N

ไปเปรียบเทียบทางโหนดที่คำหลักที่ 1 ชี้ไป

1.2 ถ้า ITEM > หรือ = คำหลักตัวที่ 2 แต่ < คำหลักตัวที่ 3 ของโหนด N

ไปเปรียบเทียบทางโหนดที่คำหลักที่ 2 ชี้ไป

1.3 ถ้า ITEM > หรือ = คำหลักตัวที่ 3 แต่ < คำหลักตัวที่ 4 ของโหนด N
ไปเปรียบเทียบทางโหนดที่คำหลักที่ 3 ชี้ไป

1.4 ถ้า ITEM > หรือ = คำหลักตัวที่ 4 แต่ < คำหลักตัวที่ 5 ของโหนด N
ไปเปรียบเทียบทางโหนดที่คำหลักที่ 4 ชี้ไป

1.5 ถ้า ITEM > หรือ = คำหลักตัวที่ 5 แต่ < คำหลักตัวที่ 6 ของโหนด N
ไปเปรียบเทียบทางโหนดที่คำหลักที่ 5 ชี้ไป

1.6 ถ้า ITEM > หรือ = คำหลักตัวที่ 6 แต่ < คำหลักตัวที่ 7 ของโหนด N
ไปเปรียบเทียบทางโหนดที่คำหลักที่ 6 ชี้ไป

1.7 ถ้า ITEM > หรือ = คำหลักตัวที่ 7 แต่ < คำหลักตัวที่ 8 ของโหนด N
ไปเปรียบเทียบทางโหนดที่คำหลักที่ 7 ชี้ไป

1.8 ถ้า ITEM > หรือ = คำหลักตัวที่ 8 แต่ < คำหลักตัวที่ 9 ของโหนด N
ไปเปรียบเทียบทางโหนดที่คำหลักที่ 8 ชี้ไป

1.9 ถ้า ITEM > หรือ = คำหลักตัวที่ 9 แต่ < คำหลักตัวที่ 10 ของโหนด N
ไปเปรียบเทียบทางโหนดที่คำหลักที่ 9 ชี้ไป

1.10 ถ้า ITEM > หรือ = คำหลักตัวที่ 10 ของโหนด N
ไปเปรียบเทียบทางโหนดที่คำหลักที่ 10 ชี้ไป

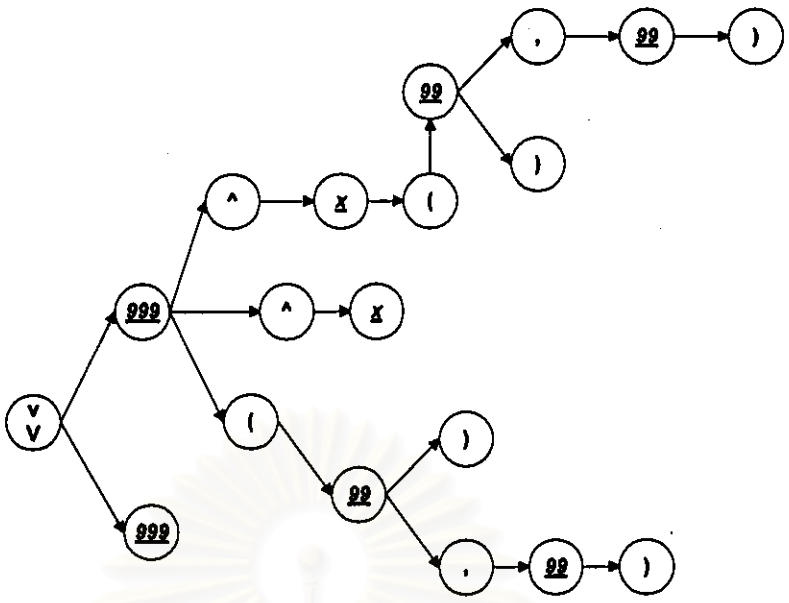
2. ทำซ้ำขั้นตอนที่ 1. โดยเปลี่ยนโหนด N เป็นโหนดใหม่ที่ถูกชี้ไป จนปรากฏเงื่อนไขดังนี้

2.1 พบโหนด N ที่มีค่า ITEM อยู่ ในกรณีที่การค้นหาประสบผลสำเร็จ

2.2 พบทรีย่อยว่าง หมายถึงการค้นหาไม่พบข้อมูล

2.8 <u>ทรัย</u> (Trie)

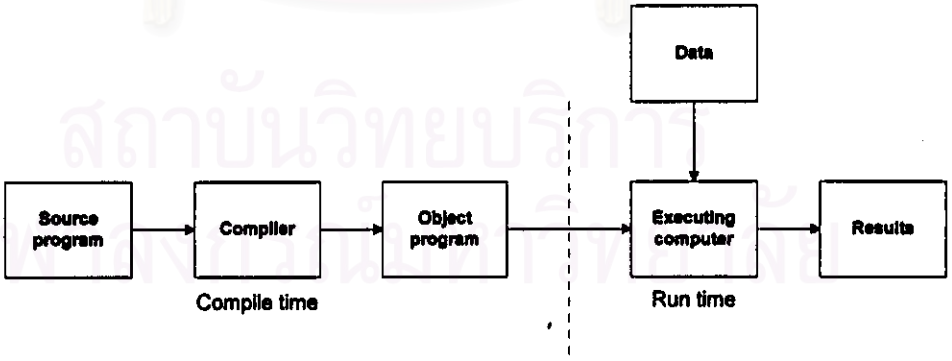
(Kruse, 1991) ทรัย คือการบรรยายลักษณะของคำ โดยพิจารณาคำเป็นลำดับของอักขระ (ตัวอักษร หรือตัวเลข) และใช้อักขระเหล่านี้ แยกออกเป็นกิ่งแขนงเส้นทางของสายอักขระแต่ละระดับ สามารถแสดงในรูปของต้นไม้ (tree) ได้ เมื่อทำการป้อนสายอักขระเข้าไปและทำการตรวจสอบจากทรัย จะสามารถบอกได้ว่าสายอักขระนั้น เป็นคำที่ถูกต้องตามข้อกำหนดหรือเป็นคำในพจนานุกรมที่กำหนดหรือไม่



รูปที่ 2.12 แสดงตัวอย่างของทรัยในการวิเคราะห์รูปแบบการแสดงผลชนิดหนึ่ง

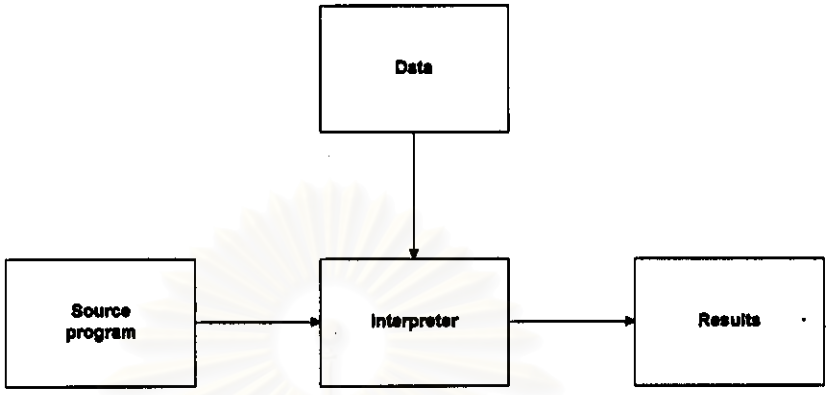
2.9 ตัวแปลโปรแกรม

(Trembley, 1985) ตัวแปลโปรแกรม คือ โปรแกรมที่รับเอาโปรแกรมดิบ (Source program) เข้าไปแล้วทำการแปลเป็นโปรแกรมผล (Object program หรือ Target program) ถ้าโปรแกรมดิบเป็นภาษาแอสเซมบลีและโปรแกรมผลที่ได้รับเป็นภาษาเครื่อง เรียกตัวแปลโปรแกรมนั้นว่า แอสเซมเบลอร์ ส่วนตัวแปลที่ทำหน้าที่แปลภาษาระดับสูง เช่น ภาษาฟอร์แทรน โคบอล ปาสคาล ฯลฯ ให้เป็นภาษาเครื่องหรือแปลเป็นภาษาแอสเซมบลี เรียกว่า คอมไพเลอร์ (Compiler)



รูปที่ 2.13 กระบวนการแปลโปรแกรมของตัวแปลโปรแกรม (Compilation Process)

ตัวแปลอีกประเภทหนึ่ง เรียกว่าตัวแปลคำสั่ง (Interpreter) ซึ่งจะทำการแปลโปรแกรมดิบ และอ่านข้อมูลเข้ามาทำการประมวลผลในเวลาเดียวกัน ดังนั้นการแปลความหมายของโปรแกรมดิบจะเกิดในช่วงเวลาการประมวลผล (Run Time) โดยไม่มีการสร้างโปรแกรมผลเก็บไว้ก่อนการประมวลผล

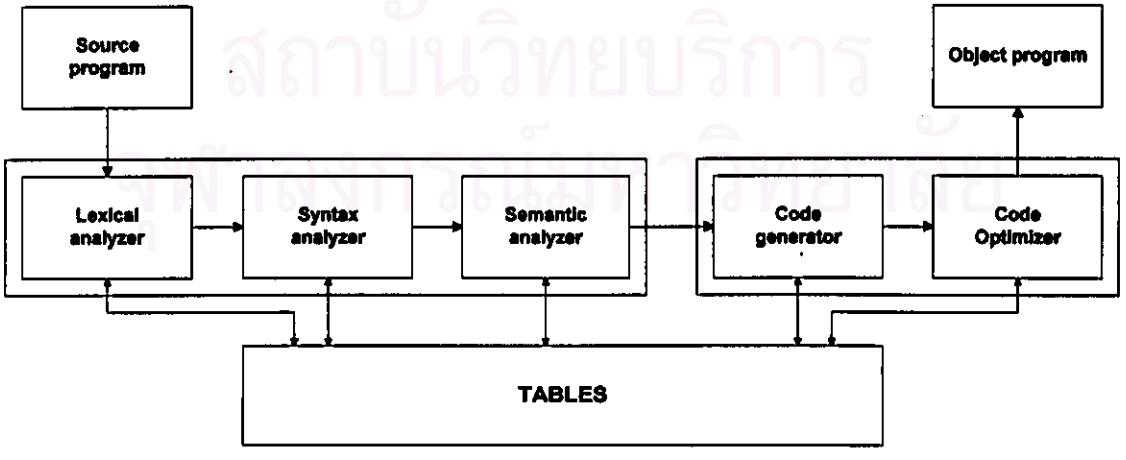


รูปที่ 2.14 กระบวนการแปลโปรแกรม ของตัวแปลคำสั่ง (Interpretation Process)

ในการใช้ไมโครคอมพิวเตอร์โดยทั่วไป จะพบว่าตัวแปลคำสั่งที่นิยมใช้กันอย่างกว้างขวางมากกว่าตัวแปลโปรแกรม เหตุผลอย่างหนึ่งก็คือ การใช้ตัวแปลคำสั่งสามารถตอบโต้กับผู้ใช้ได้ทันทีในการออกคำสั่ง แต่ละคำสั่ง โดยไม่จำเป็นต้องสร้างชุดคำสั่งให้ครบแล้วจึงแปลโดยตัวแปลโปรแกรม ซึ่งต้องทำขั้นตอนหลายขั้นตอน ตัวอย่างได้แก่ตัวแปลคำสั่งภาษาเบสิก

<u>รูปแบบทั่วไปของตัวแปลโปรแกรม</u>

รูปแบบพื้นฐานของตัวแปลโปรแกรมสามารถแสดงได้ดังรูปต่อไปนี้



รูปที่ 2.15 รูปแบบพื้นฐานของตัวแปลโปรแกรม

กระบวนการวิเคราะห์ศัพท์ (Lexical Analysis)

ขั้นตอนนี้ทำหน้าที่อ่านอักขระจากโปรแกรมต้นแบบ ซึ่งจะประกอบขึ้นมาจากสายอักขระ (String) ของสัญญลักษณ์ต่างๆ อันได้แก่ ตัวอักษร, ตัวเลข, อักขระพิเศษ ซึ่งส่วนย่อยเหล่านี้จะประกอบขึ้นเป็นองค์ประกอบของโปรแกรม ได้แก่ นิพจน์ทางคณิตศาสตร์, ตัวแปร, ค่าคงที่, คำสงวนต่างๆ มาทำการแยกย่อยออกเป็นหน่วยย่อยหรือโทเค็น (Token) เบื้องต้น แล้วจึงนำหน่วยย่อยเหล่านี้ไปใช้ในขั้นตอนต่อไป

กระบวนการวิเคราะห์วากยสัมพันธ์ (Syntactical Analysis)

ขั้นตอนนี้ทำหน้าที่ต่อจากตัววิเคราะห์ศัพท์ โดยจะทำการตรวจสอบและวิเคราะห์ความถูกต้องของไวยากรณ์ภาษา ซึ่งทำหน้าที่โดยส่วนของตัวแปลโปรแกรมที่เรียกว่า ตัววิเคราะห์กระจาย (Parser)

ส่วนนำเข้า (Input) ของตัววิเคราะห์กระจาย ก็คือ ผลลัพธ์ที่ได้จากการทำงานของ กระบวนการวิเคราะห์ศัพท์ หรือ โทเค็น เมื่อผ่านการตรวจสอบและวิเคราะห์ความถูกต้องของไวยากรณ์แล้ว ผลลัพธ์จะอยู่ในรูปแบบต่างๆ แต่ที่นิยมกันมากจะสร้างในแบบที่เรียกว่า ต้นไม้วิเคราะห์กระจาย (Parse Tree)

กระบวนการวิเคราะห์อรรถศาสตร์ (Semantic Analysis)

ขั้นตอนนี้ทำหน้าที่ตรวจสอบความหมายของภาษาและสร้างรหัสของชุดคำสั่งที่ผ่านการตรวจสอบความถูกต้องของโครงสร้างแล้ว โดยทำการเปลี่ยนแปลงต้นไม้วิเคราะห์กระจาย ให้อยู่ในรูปแบบการทำงานที่ใกล้เคียงกับภาษาเครื่อง ซึ่งรหัสดังกล่าวเรียกว่า รหัสชั่วคราว (Intermediate Code Generation)

การวิเคราะห์และสังเคราะห์รหัสที่เหมาะสม (Code Optimizer)

เป็นขั้นตอนที่ทำการแก้ไขปรับปรุงรหัสชั่วคราวที่ได้จากกระบวนการวิเคราะห์อรรถศาสตร์ ให้มีประสิทธิภาพในการทำงานมากขึ้นซึ่งตัวแปลโปรแกรมบางตัวอาจจะไม่มีขั้นตอนนี้ก็ได้

การสร้างรหัสเครื่อง (Code Generation)

ขั้นตอนนี้เป็นการเปลี่ยนรหัสชั่วคราวเป็นคำสั่งรหัสเครื่องที่สามารถทำงานได้จริงๆโดยวิธีการถอด (Map) จากรหัสชั่วคราวออกมาเป็น รหัสภาษาเครื่องโดยตรง

2.10 ตัวประสานผู้ใช้แบบกราฟิก (Graphical user interface)

ในอดีตการใช้งานคอมพิวเตอร์มีข้อจำกัดเฉพาะตัวอักษรหรือข้อความ ตัวอักษรแต่ละตัวจะถูกส่งผ่านอุปกรณ์รับข้อมูลเข้าไปยังหน่วยประมวลผลหรือหน่วยความจำ แล้วส่งไปแสดงผลทางจอภาพ ในยุคเริ่มต้นข้อมูลเข้าที่ป้อนเข้าสู่ระบบมีลักษณะเป็นแบบกลุ่ม (batch) คือ ต้องบันทึกข้อมูลลงกระดาษเจาะรูก่อนแล้วส่งเข้าเครื่องเพื่อประมวลผลต่อไป ต่อมาระบบเริ่มมีการเชื่อมต่อกับผู้ใช้มากขึ้น มีการทำงานในลักษณะเชื่อมตรง (on-line) การใช้เป็นพิมพ์และจอภาพจึงแพร่หลายและเป็นที่นิยม การเชื่อมต่อระหว่างผู้ใช้กับคอมพิวเตอร์หรือที่เรียกว่า ระบบตัวประสานผู้ใช้ ก็ได้พัฒนาให้ก้าวหน้าขึ้นเป็นลำดับ ระบบเชื่อมตรงหรือที่เรียกว่า ระบบเชิงโต้ตอบ (interaction) ทำให้มีการใช้เมนูหรือรายการเลือก และมีการพัฒนาเมนูในรูปแบบเป็นแถบสว่างให้เลือก จนในที่สุดมีการใช้หลักการซ่องหน้าต่างโดยใช้เมนูแบบผุดขึ้น (pop-up) และดึงลง (pull-down)

(Peddie, 1992) ในปี ค.ศ. 1970 บริษัท ซีรอกซ์ (Xerox) ได้สร้างทีมงานวิจัยขึ้นมาชื่อว่า Palo Alto Research Center หรือ PARC ได้ออกแบบระบบตัวประสานผู้ใช้ขึ้นใหม่เป็นแบบกราฟิก โดยนำไปใช้กับเครื่อง Star ต่อมา สตีฟ จอปส์ (Steve Jobs) ได้นำไปใช้กับเครื่องแอปเปิ้ลแมคอินทอช จนประสบความสำเร็จอย่างสูง ทำให้เครื่องแมคอินทอชเป็นที่รู้จักอย่างกว้างขวางเพราะมีตัวประสานผู้ใช้ที่สวยงามน่าใช้ สามารถเรียนรู้การใช้งานได้ง่ายและใช้งานได้สะดวก เพียงแค่ใช้เมาส์ในการเลือกสัญรูป (icon) ซึ่งเป็นรูปภาพเล็กๆในรูปต่างๆ อันเป็นตัวแทนของแฟ้มข้อมูลหรือโปรแกรมก็สามารถเรียกใช้แฟ้มข้อมูลหรือโปรแกรมนั้นๆได้ และยังใช้เมาส์ในการเลือกเมนู เลือกหน้าต่าง (window) ซึ่งมีได้หลายๆหน้าต่าง ทำให้ผู้ใช้รู้สึกใช้งานได้ง่าย

หลังจากนั้นต่อมาตัวประสานผู้ใช้แบบกราฟิก จึงเป็นที่ยอมรับอย่างกว้างขวาง เช่น โอเอสทู (OS/2) ของบริษัทไอบีเอ็ม ไมโครซอฟท์วินโดวส์ (Microsoft Windows) ของบริษัทไมโครซอฟท์ เป็นต้น

ไมโครซอฟท์วินโดวส์

(จิรพัฒน์ จันทร์เจิดศักดิ์, 2536) ตัวประสานผู้ใช้แบบกราฟิกในท้องตลาดนั้นมีอยู่มากมายหลายตัว แต่ตัวที่นิยมใช้งานกันอย่างแพร่หลายบนเครื่องไมโครคอมพิวเตอร์คือ ไมโครซอฟท์วินโดวส์ของบริษัทไมโครซอฟท์ ซึ่งมีการพัฒนาอย่างต่อเนื่อง โดยรุ่นที่มีชื่อเสียงมากคือ ไมโครซอฟท์วินโดวส์ รุ่น 3.1 , รุ่น 3.11 และรุ่น 95

ระบบไมโครซอฟท์วินโดวส์ นั้นมีคุณสมบัติสำคัญ คือ

1. มีระบบตัวประสานผู้ใช้แบบกราฟิก ประกอบด้วยหน้าต่าง, เมนู, กรอบโต้ตอบ (dialog box) และตัวควบคุม (control) ชนิดต่างๆ

2. มีระบบรับข้อมูลเข้าแบบใหม่ที่สะดวกกว่าเดิม ได้แก่กิจกรรมต่างๆ เช่น การชี้ (point), การคลิก (click), การกด (press), การลาก (drag), การคลิกคลิก (double click), การเลื่อนรายการ (scroll), การป้อน (key)

3. มีระบบจัดการงานแบบหลายภารกิจ (multitasking)

4. มีการเชื่อมโยงและการใช้ข้อมูลร่วมกันระหว่างโปรแกรมประยุกต์

ข้อแตกต่างของโปรแกรมบนดอสและโปรแกรมบนวินโดวส์

1.ระบบตัวประสานผู้ใช้

จุดประสงค์หลักอันหนึ่งของการออกแบบวินโดวส์ ก็เพื่อที่จะให้ผู้ใช้ สามารถใช้งานโปรแกรมประยุกต์ หลายๆ โปรแกรมประยุกต์พร้อมกัน รวมทั้งยังให้ผู้ใช้สามารถมองเห็นการทำงานของทุกๆโปรแกรมประยุกต์จึงต้องแบ่งปันจอภาพ เพื่อแสดงผลต่อผู้ใช้ ใหญ่บ้าง เล็กบ้าง ตามแต่ผู้ใช้ต้องการและสามารถแสดงผลให้แก่ผู้ใช้ได้ตลอดเวลา นอกจากเป็นอุปกรณ์แสดงผลขั้นพื้นฐานของโปรแกรมประยุกต์แล้ว ยังเป็นสิ่งที่ผู้ใช้สามารถติดต่อกับโปรแกรมประยุกต์ได้โดยผ่านส่วนประกอบต่างๆของวินโดว์ เช่น เมนู ปุ่มบังคับ แท่งเลื่อนรายการ เป็นต้น

2.ระบบการรับข้อมูลเข้า

โปรแกรมที่ทำงานบนดอส เมื่อจะรับข้อมูลจากผู้ใช้ จะต้องรอจนกว่าผู้ใช้จะใส่ข้อมูลจริงๆเสียก่อน หากในระหว่างนั้นผู้ใช้ทำการใส่ข้อมูลทางอื่น เช่น เมาส์ ก็จะไม่สามารถรับข้อมูลเข้ามาได้

โปรแกรมที่ทำงานบนวินโดวส์ จะมีลักษณะการทำงานอย่างหนึ่งที่เรียกว่า ตอบสนองข่าว (message-driven) หรือ ตอบสนองเหตุการณ์ (event-driven) คือโปรแกรมจะทำงานโดยอาศัยข่าว (message) การที่มีข่าวเกิดขึ้นก็คือการที่เกิดเหตุการณ์ ขึ้นในระบบ เช่น มีการกดปุ่มแป้นพิมพ์ การกดปุ่มเมาส์ ซึ่งวินโดวส์จะรวบรวมข้อมูลเข้าจากทุกอุปกรณ์แล้วเก็บไว้ใน แถวคอยข่าว (message queue) เมื่อโปรแกรมประยุกต์ต้องการอ่านข้อมูลเข้าก็เพียงแต่อ่านจากแถวคอยข่าวเท่านั้น โปรแกรมประยุกต์จะได้รับข้อมูลจากผู้ใช้โดยไม่มีการสูญหาย

3.ระบบกราฟิก

ดอสไม่มีการสนับสนุนระบบกราฟิก ในขณะที่วินโดวส์มีการสนับสนุนระบบกราฟิกอย่างมีประสิทธิภาพ นอกจากนี้ยังสามารถใช้ได้กับอุปกรณ์แสดงผลทุกชนิด ที่สามารถแสดงผลเป็นกราฟิกได้ สามารถใช้คำสั่งเดียวกัน รูปแบบเดียวกัน วาดรูปได้ทั้งบนจอภาพ เครื่องพิมพ์ และพล็อตเตอร์ ฯลฯ โดยวินโดว์จะอาศัยโปรแกรมขับอุปกรณ์ (device driver) ในการแสดงผลให้เหมาะสมกับอุปกรณ์ที่ใช้แต่ละแบบ

4.ระบบจัดการงาน

โปรแกรมที่ทำงานบนดอส โดยทั่วไปจะสมมุติตัวเองว่าสามารถเข้าไปครอบครองทรัพยากรต่างๆของระบบทั้งอุปกรณ์ด้าน รับข้อมูลเข้า, ส่งข้อมูลออก, หน่วยความจำ และหน่วยประมวลผล ในลักษณะเป็นงานเดียวของระบบ เมื่อโปรแกรมบนดอสรอรับการกดแป้นพิมพ์ จะไม่สามารถไปทำงานอย่างอื่นได้

โปรแกรมที่ทำงานบนวินโดวส์ หรือระบบตัวประสานผู้ใช้แบบกราฟิกอื่นๆ จะต้องมีการแบ่งปันทรัพยากรเหล่านั้นร่วมกัน คือเมื่อโปรแกรมใช้ทรัพยากรจนเสร็จเรียบร้อยแล้ว ต้องคืนทรัพยากรกลับคืนไปให้ระบบ ซึ่งมีวิธีเริ่มใช้ทรัพยากรและคืนทรัพยากรให้ระบบโดยอาศัยข่าว

2.11 โปรแกรมเดลไฟ

เป็นซอฟท์แวร์ ที่มีความสามารถในการพัฒนาโปรแกรมประยุกต์บนระบบปฏิบัติการวินโดวส์ โดยรวมเอาความสามารถของภาษาปาสคาล และการเขียนโปรแกรมแบบเชิงวัตถุ (object-oriented programming) เข้าไว้กับเครื่องมือการออกแบบแบบภาพ (visual design tools) ทำให้สามารถเขียนโปรแกรมในลักษณะภาพ (visual) โดยเริ่มต้นที่การออกแบบส่วนตัวประสานผู้ใช้ ด้วยการนำเอาส่วนประกอบย่อย หรือ ที่โปรแกรมเดลไฟเรียกว่า ออบเจ็กต์ (object) หรือคอมโปเนนต์ (component) จากส่วนคอมโปเนนต์แพลเลต (component palette) มาประกอบกันเป็นหน้าตาของโปรแกรมที่ต้องการก่อนทำการกำหนดคุณสมบัติของแต่ละออบเจ็กต์นั้น กำหนดเหตุการณ์ ที่จะเกิดกับออบเจ็กต์นั้น แล้วจึงเขียนโปรแกรมตอบสนองต่อเหตุการณ์แต่ละเหตุการณ์ หรือที่เรียกว่า การเขียนโปรแกรมแบบตอบสนองเหตุการณ์ (event-driven programming) ซึ่งเป็นรูปแบบการเขียนโปรแกรมที่เหมาะกับระบบตัวประสานผู้ใช้แบบกราฟิก (Graphical User Interface, GUI)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย