

# 온라인 情報檢索(2)

柳 慈 卿\*

## 4. 온라인 情報檢索시스템

온라인을 利用한 書誌情報檢索시스템의 數는 약 200個에 달한다. 여기에서는 現在 많이 利用되고 있는 온라인 情報檢索시스템들의 歷史, 使用機種, 檢索例 및 각 시스템이 取扱하고 있는

데이터베이스들에 대해 간단히 說明하고자 한다.

### 4.1 BASIS-70

BASIS (Battelle Automated Search Information System)는 Battelle Memorial Institute

```

BASIS 70
(PHASE 1)
IS ON LINE↓
DO YOU DESIRE OPERATING INSTRUCTIONS?
TYPE YES OR NO/
NO^
PLEASE ENTER YOUR LAST NAME. /
PENNIMAN^
ENTER THE NAME OF DATA BASE TO BE SEARCHED. /
CCIC^
ENTER YOUR SEARCH ONE TERM AT A TIME. /
1/
FISH^
135 ITEMS^
2/
AQUATIC LIFE^
144 ITEMS^
3/
    
```

(a)

```

GULF OF ALASKA^
35 ITEMS^
4/
(LIST ALL)^
ITEMS-NO. -LINE---
*** 135 1/FISH
*** 144 2/AQUATIC LIFE
*** 35 3/GULF OF ALASKA
ENTER YOUR REQUEST /
4/
(1 OR 2 AND 3)^
17 ITEMS^
ENTER YOUR REQUEST /
5/
    
```

(b)

```

(DISPLAY 4)^
THE DATA ELEMENTS FOR THE CCIC DATA BASE ARE:
1-ACCESSION NUMBER,
2-AUTHOR(S),
3-TITLE,
4-BIBLIOGRAPHIC DATA,
5-INDEX TERMS,
6-ABSTRACT.
WHAT FIELDS DO YOU WANT TO SEE?
ENTER FIELD NUMBERS SEPARATED BY COMMAS OR ALL /
3,2,4,6^
THIS PRINTOUT COULD BE LENGTHY.
HOW MANY DO YOU WANT TO SEE FIRST? /
1
    
```

(c)

```

ITEMS FROM THE CCIC DATA BASE ARE .....
ITEM 1
ACCESSION NUMBER IS: 4425
TITLE IS: MIDWATER TRAWLING STUDIES IN THE NORTH PACIFIC.
AUTHOR(S): WILLIAM ARON
BIBLIOGRAPHIC DATA: LIMNOLOGY AND OCEANOGRAPHY, VOL. 4/4, PP 409-418, 1959, ILLUSTRATIONS, MAPS, TABS, 11 REFS.
ABSTRACT: A MODIFIED VERSION OF THE ISAACS-KIDD MIDWATER TRAWL WAS USED FOR 149 EXPLORATORY HAULS IN THE NORTHWESTERN PACIFIC AT DEPTHS RANGING FROM 20 TO 250 METERS. CATCHES ARE DISCUSSED IN TERMS OF SUCH FACTORS AS DIURNAL MIGRATION, SEASONAL VARIATION, PATCHINESS, AND GEOGRAPHIC VARIATION. THERE WAS WIDE DIF
    
```

(d)

\*KORSTIC 電子計算室

그림 1. BASIS의 檢索例

의 Columbus Laboratory에서 개발한 시스템으로 CDC 6400 컴퓨터를 사용한다. 이 시스템은 60명의 터미널 利用者에게 同時に 서어비스를 提供할 수 있으며 부울린(boolean) 形態의 檢索式에 의하여 記事本文이나 索引된 데이터베이스에 대하여 探索을 행한다.

1971년에 약 30個의 데이터 파일을 處理하던 이 시스템은 후에 큰 데이터베이스들을 더 追加하였다. NTIS와 CAC 등이 이 때 包含되었으며 이들은 TYMSHARE 네트워크에 의하여 連結되어 있다.

BASIS-70에서는 각 檢索用語를 直接 부울린으로 連結시킬 수는 없고, 하나씩 불러낸 다음 후에 이것을 論理的으로 連結시킨다. CCIC (Cold Climate Information Center)의 抄錄파일에서 願하는 데이터를 찾는 例가 그림 1에 나타나 있다.

4.2 BOLD

BOLD(Bibliographic On-Line Display)란 SDC(Systems Development Corporation)에 의해 實驗적으로 만들어진 시스템으로 SDC의 IBM Q-32V 컴퓨터에 TTY, CRT 디스플레이, light pen 등을 連結, 使用토록 設計되었다. BOLD의 實驗은 U. S. Defence Documentation Center의 抄錄 데이터베이스에 대해 이루어졌다. 이 시스템에서는 어떤 形態의 부울린 檢索式도 使用 可能하며 한 用語에 關連된 關連用語 및 同義語, 綴字가 비슷한 用語들을 디스플레이에 비쳐 낼 수 있어 檢索을 돕고 있다. 利用者는 search mode와 browse mode 중 하나를 選擇할 수 있는데 search mode는 檢索에 使用된 각 用語에 대하여, 檢索된 文獻들이 어떤 用語에 의하여 檢索되었는가 하는 것을 보여준다. 表 1은 search mode의 出力例인데 여기에서 AD-284 259는 다섯개의 檢索用語로, AD-272572는 3개의, AD-274816은 2개의 檢索用語로 각각 索引되어 있음을 나타내고 있다.

Browse mode는 書誌데이터들이 어떤 主題分野로 構成되어 있는가 하는 것을 보여주는 것으로 表 2에 그 出力例가 있다. 檢索者는 자신이

表 1. BOLD Search mode의 出力例

32 ENTRIES ALL SEARCHED	1 DOCUMENTATION	2 INFORMATION RETRIEVAL	3 LIBRARY SCIENCE	4 INDEXES	5 CLASSIFICATION	6 VOCABULARY
AD - 284 259	x	x	x	x	x	
AD - 272 572	x		x	x		
AD - 278 141		x			x	x
AD - 283 335	x	x				
AD - 275 814		x		x		
AD - 274 816		x				x
AD - 271 926	x					
AD - 276 533	x					
AD - 286 511	x					
AD - 285 510	x					
AD - 271 600	x					
AD - 274 358	x					
AD - 281 909			x			
AD - 275 826				x		
AD - 275 964				x		
AD - 276 142				x		
AD - 271 124				x		
AD - 272 068				x		
AD - 278 130				x		
AD - 275 528				x		
AD - 274 026				x		
AD - 286 816				x		
AD - 274 775						x
AD - 276 376						x
AD - 274 260						x
AD - 286 637						x
AD - 266 508						x
AD - 272 133						x
AD - 274 774						x
AD - 272 401						x
AD - 277 520						x
AD - 276 066						x

表 2. BOLD Browse mode의 出力例

THE FOLLOWING CATEGORIES ARE AVAILABLE.

- DIV 1 AERONAUTICS
- DIV 2 ASTRONOMY GEOPHYSICS GEOGRAPHY
- DIV 3 CHEMICAL WARFARE
- DIV 4 CHEMISTRY
- DIV 5 COMMUNICATIONS
- DIV 6 DETECTION
- DIV 7 ELECTRICAL EQUIPMENT
- DIV 8 ELECTRONICS
- DIV 9 FLUID MECHANICS
- DIV 10 FUELS AND COMBUSTION
- DIV 11 GROUND TRANSPORTATION EQUIPMENT
- DIV 12 GUIDED MISSILES
- DIV 13 CONSTRUCTION
- DIV 14 MATERIALS (NON METALLIC)
- DIV 15 MATHEMATICS
- DIV 16 MEDICAL SCIENCES
- DIV 17 METALLURGY
- DIV 18 MILITARY SCIENCES
- DIV 19 NAVIGATION
- DIV 20 PHYSICS CHEMISTRY (NUCLEAR)
- DIV 21 PROPULSION (NUCLEAR)
- DIV 22 ORDNANCE
- DIV 23 PERSONNEL
- DIV 24 PHOTOGRAPHY
- DIV 25 PHYSICS
- DIV 26 PRODUCTION. MANAGEMENT
- DIV 27 PROPULSION SYSTEMS
- DIV 28 PSYCHOLOGY
- DIV 29 QUARTERMASTER
- DIV 30 RESEARCH
- DIV 31 SHIPS. MARINE EQUIPMENT
- DIV 32 ARTS SCIENCES (MISC)

願하는 分野를 그 중에서 選擇하여 light pen으로 指摘하면 시스템은 그 分野內의 細分類를, 그 細分類가 包含하고 있는 文獻件數와 함께 디스플레이 畫面에 보여준다.

4.3 BROWSER

BROWSER(Browsing On-Line with Selective Retrieval)은 IBM Federal Systems Division의 Williams에 의해 개발된 시스템으로 文獻本文이나 抄錄本文에 대해 檢索이 행해진다. 데이터베이스에 入力되는 文獻들은 前置詞, 冠詞들과 같은 單語를 除外한 나머지 單語가 미리 準備된 語根辭典(dictionary of word root)과 對照된 후 이에 對應되는 語根이 있으면 그 單語는 그 單語가 사용된 文獻番號에 대하여 inverted file이 만들어진다. 이 시스템에 包含된 語根辭典은 데이터베이스內에서 각 語根의 使用頻度統計 및 使用頻도에 反比例하는 Information Value(I-Value)를 包含하는데 I-Value는  $2^{-n}$

의 比重으로 表示된다. 그러므로 I-Value가 1.0이라는 것은 抄錄들의 1/2에서 그 語根을 사용하고 있음을 말하며 2.0이란 抄錄들의 약 1/4이 그 語根을 사용하고 있음을 말한다. 따라서 頻도가 낮게 사용되는 單語일수록 높은 I-Value를 갖는다.

BROWSER의 利用者は 부울린공식에 의하기 보다는 日常 使用하는 語句로서 그의 興味主題를 表現한다. 이 시스템은 IBM 360에 連結된 2260디스플레이 터미널을 使用하여 온라인 對話方式의 檢索을 가능토록 하며 表 3과 表 4에서 볼 수 있는 것과 같이, 利用者の 檢索要求가 分析된 後 생긴 語根의 使用頻도와 I-Value 및 그 語根을 使用한 抄錄番號를 보여준다.

BROWSER는 25,000件的 獨逸 特許 抄錄, 9,000件的 英國特許抄錄, 8,000件的 美國 Defence Documentation의 抄錄과 Navy Automated Research and Development Information System(NARDIS)의 1,600件的 抄錄을 對象으로 始作되었으나, 1970년 獨逸 IBM의 特

表 3. BROWSER에 의해 分析된 檢索要求의 出力例

I am interested in the automatic classification of documents into subject categories, groups, clusters, or clumps, using discriminant or latent factors. Other statistical indexing techniques include association matrices or correlation coefficients based on word occurrences.								
ASSOC	440	4.2	AUTOMAT	479	4.1	CATEGOR	123	6.0
CLASSIF	216	5.2	CLUMP	7	10.2	CLUSTER	28	8.2
COEFF	328	4.6	CORRELAT	308	4.7	DISCRIM	77	6.7
DOCUMENT	263	4.9	FACTOR	503	4.0	GROUP	336	4.6
INDEX	268	4.9	INCLU	1376	2.6	INTEREST	336	4.6
LATENT	11	9.5	MATRI	219	5.2	OCCUR	149	5.8
STATIST	316	4.7	SUBJECT	484	4.1	TECHNIC	359	4.5
WORD	171	5.6						

表 4. BROWSE에 의한 Response Index

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
AUTOMAT	A			A	A	A	A			A	A			A			A	A	A					A	
CLASSIF	C	C		C		C	C	C			C	C			C	C								C	
DOCUMENT	D	D			D	D	D	D	D	D	D	D	D					D							
DISCRIM								D							D										D
STATIST			S		S				S		S		S	S						S		S	S		S
FACTOR	F		F	F					F							F									F
LATENT	L																								
WORD			W				W	W			W			W	W	W	W							W	W

許部に 採用된 以後 現在 약 100,000 件의 英獨 特許抄錄을 取扱한다.

### 4.4 Data Central

Data Central은 Mead Corporation의 從屬 會社인 Mead Central에 의해 개발되어 販賣되고 있는 多目的 文獻檢索시스템이다. 데이터베이스에 入力되는 文獻本文은 stop list에 包含된 單語를 除外한 모든 單語에 대하여 inverted file 이 만들어 진다. 부울린公式에 의한 檢索과 單語의 斷切(truncation)도 使用可能하며 KWIC 索引을 디스플레이에 보여 주므로서 利用者の 檢索을 돕는다. IBM 360 과 370에 맞도록 設計되었으나 어느 컴퓨터에로도 쉽게 變換可能하며 터미널로는 TTY, IBM1050 터미널과 여러 종류의 CRT디스플레이가 連結 可能하다.

Data Central은 數個의 檢索시스템에 採用되어 있으며, 그 중 重要的 것으로는 Ohio State Bar Association의 法律關係 데이터베이스에 檢索시스템인 OBAR와 National Institute of Neurological Disease and Stroke의 Epilepsy Abstract Retrieval System(EARS)을 들 수

있다. 또한 美國 心理學會에서는 이 시스템을 使用하여 1967년 이후의 Psychological Abstracts 에 대한 檢索서비스를 提供하고 있으며 이것을 PADAT(Psychological Abstracts Direct Access Terminal)라 부른다.

### 4.5 DIALOG

DIALOG는 Lockheed Palo Alto Research Laboratory에 의해 개발된 專用시스템으로 NASA에서 保有하고 있는 약 300,000 件의 記事를 對象으로 設計되었다. 1966년 처음 使用된 이 시스템은 1969년 DIALOG II란 이름으로 改定되었으며, 현재에도 계속 改定 補完되고 있는 시스템이다. 使用하고 있는 機種은 IBM 360/30(32KB Core)으로 프로그램을 위한 두개의 IBM 2311 disc pack(7.5million byte/each)과 데이터베이스의 貯藏 및 檢索을 위한 IBM 2321 Data cell, IBM 1443프린터 등이 함께 使用되고 있다. 이외에 터미널로는 IBM 2260 디스플레이 및 1053 프린터 터미널이 使用되고 있다. 요즘에는 이것을 IBM 360/40과 이에 連結된 CRT 콘솔을 使用하는데 DIALOG의 콘솔은 그림 2와 같이 여

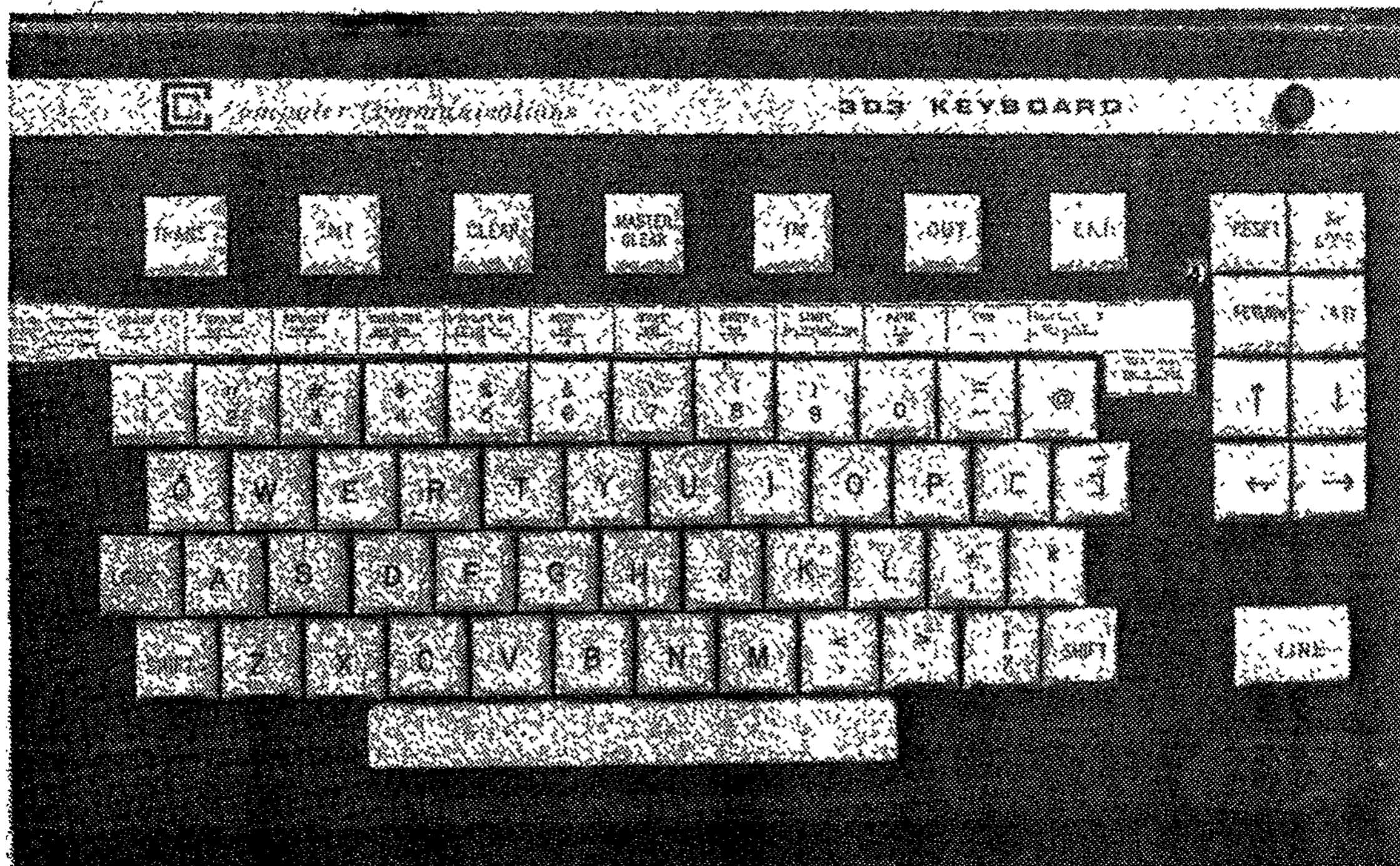


그림 2. DIALOG 콘솔

러 命令을 위한 特定鍵盤을 갖고 있으며 그들이 意味하는 명령은 다음과 같다.

○ EXPAND Term X : 시스템內에 包含된 디소오러스에서 X 라는 用語에 알파벳順으로 가까운 用語를 디스플레이 하라는 命令이다. 각 用語는 이들에 臨時로 붙여진 ID番號, 시스템內에서 그 用語로 索引된 回數가 그 用語에 關連된 用語의 數와 함께 디스플레이 된다.

○ SELECT : EXPAND 命令에 의해 提示되었던 用語들 중에서 앞으로의 檢索過程에서 使用할 用語를 指定하는 命令으로 用語의 ID番號를 使用하여 指定한다. 여기에서 볼 수 있듯이 DIALOG는 檢索用語가 디소오러스에 局限된 시스템이다.

○ COMBINE : SELECT에 의해 選擇된 用語들을 組合하는 命令으로 부울린公式을 使用한다.

例) COMBINE A+B (A or B or both)

COMBINE A\*B (A and B)

COMBINE A-B (A but not B)

COMBINE 命令의 結果로는 檢索論理를 滿足시키는 文獻의 數가 디스플레이 된다.

○ DISPLAY : 檢索된 文獻의 書誌的 事項을 CRT 디스플레이에 비추어 내라는 命令이다. 디스플레이되는 順序는 파일內의 레코드 順序에 一致된다.

○ ITEM : 다음 레코드를 디스플레이 하라는 命令이다.

○ PAGE : 한 레코드의 길이가 한 畫面에 비추어 낼 수 있는 길이보다 긴 경우에 使用되며 機能은 ITEM과 같다.

○ KEEP : 檢索된 文獻을 後에 온라인 혹은 오프라인 프린터에 찍기 위하여 保留하여 두라는 命令이다.

○ TYPE : 選擇된 文獻을 콘솔 프린터에 찍으라는 命令이다.

○ PRINT : 選擇된 文獻 혹은 그들의 一部分을 오프라인 고속프린터로 찍으라는 命令이다.

이러한 命令들은 NASA의 RECON 시스템에서도 똑같이 使用되고 있는데 이것은 NASA의 RECON도 Lockheed에 의하여 개발되었으며 이때 DIALOG의 言語가 RECON에 吸收되었기

때문이다.

DIALOG는 NASA뿐 아니라 U. S. Office of Education의 ERIC (Educational Resources Information) 데이터베이스, U. S. Atomic Energy Commission의 Nuclear Science Abstracts 데이터베이스, NTIS의 U. S. Government Research and Development 데이터베이스, CCM Information Service에 의해 編集된 PANDEX 파일의 處理에도 適用되고 있다. 1973년에는 ERIC (115,000 記事), NTIS (150,000), PANDEX (400,000) 데이터베이스가 TYMSHARE 네트워크를 통해 Lockheed에 登錄된 利用者들에게 利用되게 되었다.

#### 4.6 INTREX

INTREX (Information Transfer Experiments)는 온라인 도서관시스템의 試圖로서 Massachusetts Institute of Technology의 研究實驗 結果 만들어진 시스템이다. IBM 7094 컴퓨터를 使用하고 있는 이 시스템은 약 30명의 터미널 利用者에게 同時에 서비스를 提供할 수 있으며 平均 應答時間은 약 15~20秒이다. 이 시스템의 構成은 그림 3과 같으며 여기에서 볼 수 있듯이 材料工學 分野의 20,000件的 記事目錄은 컴퓨터에 收錄하고 이들 記事의 本文은 마이크로피쉬에 收錄하였다.

雜誌記事 目錄은 1) IBM 2741이나 Datel 30과 같은 TTY 터미널 2) ARDS (Advanced Remote Display Station)라 불리는 CRT 디스플레이 터미널 3) INTREX를 위해 特別히 만들어진 BRISC (Buffered Remote Interactive Search Console) 디스플레이 터미널 등의 터미널을 통해 찾아볼 수 있는데 그중 BRISC는 字形이 크고 鮮明하며 한 페이지의 內容을 貯藏해 두었다가 디스플레이 할 수 있는 機能이 있기 때문에 利用者의 愛好를 받았다. 記事本文은 1/18의 크기로 縮小되어 마이크로피쉬 形態로 CARD (Compact Automatic Retrieval Display)에 貯藏되며 檢索된 記事本文의 디스플레이를 위하여는 그 特定目的을 위한 CRT 디스플레이나 ARDS 터미널이 使用된다. 한장의 fiche에는 60

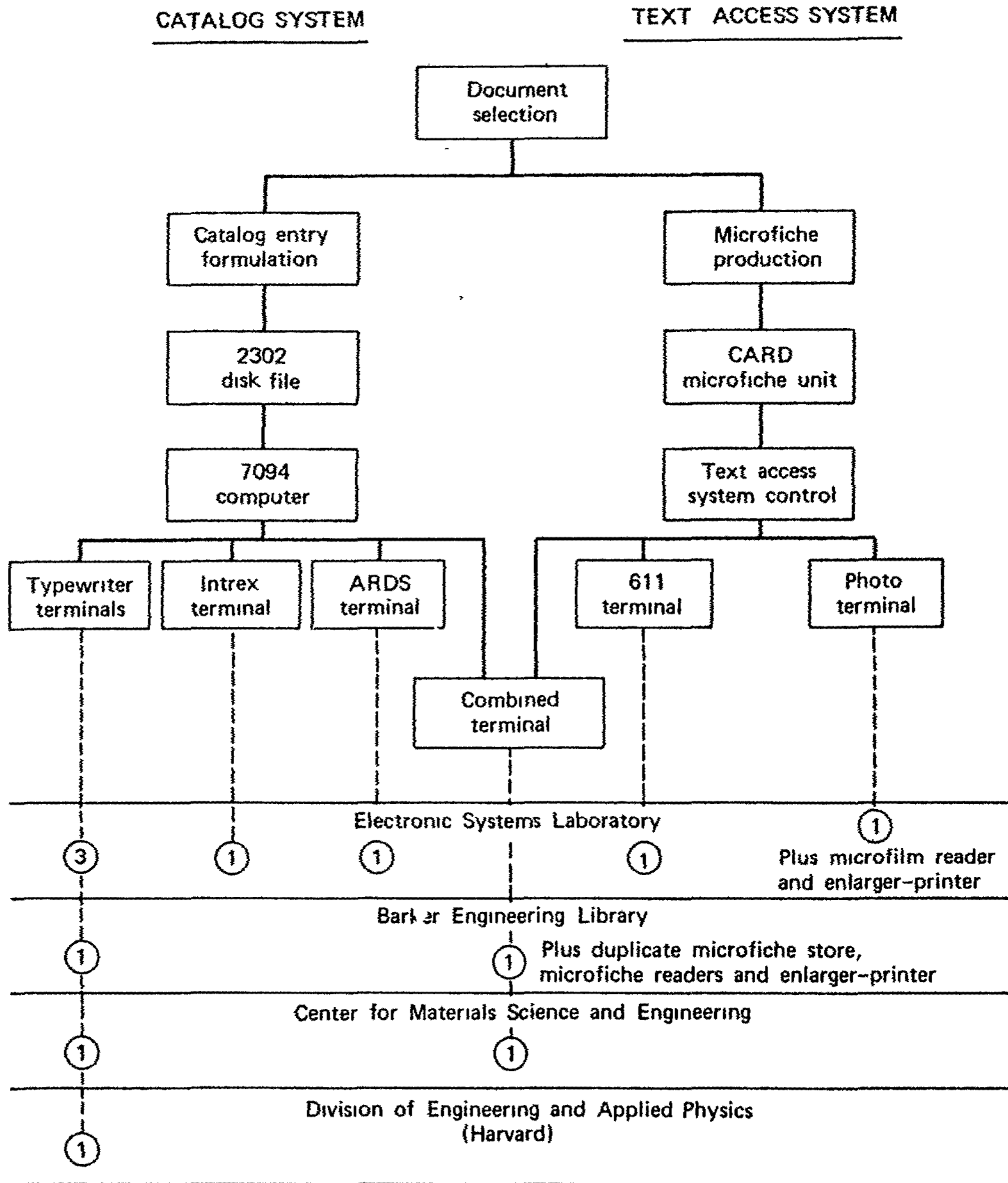


그림 3. Intrex 시스템 構成圖

페이지를 收錄하는데 한 記事가 平均 6페이지라 할 때 10개의 文獻이 하나의 fiche에 收錄된다 하겠다. 다른 文獻檢索시스템이 文獻의 書誌的 事項만을 提供하는데 比하여 INTREX는 文獻의 書誌的 事項과 함께 願하는 경우에는 記事本 文을 提供하는 시스템이기 때문에 어떤 사람들은 "도서관 自体를 提供하는 시스템"이라고 評하기도 한다. 記事本文은 本文 提示의 命令後 7秒以 內에 디스플레이에 한 페이지씩(56자×31행) 나 타나며 3秒마다 그 內容이 바뀌어 다음 페이지 를 보여준다. 또한 記事本文은 마이크로필름 및 8½×11인치 종이에 프린트된 形態로도 出力可

能하다.

#### 4.7 LEADER

LEADER(Lehigh Automatic Device for Efficient Retrieval)는 1968년 Lehigh대학 情報科學센터에서 情報시스템의 모델로 만든 시스템으로 IBM 1800 Process Control 컴퓨터와 IBM 2260디스플레이를 使用하였다. 情報要求는 英語文章과 같은 形式으로 入力되며 文獻이 이 시스템內에서 處理된 것과 같은 方法으로 處理된 다음, 文獻과 要求가 對照된 後 미리 定하여진 基

準値보다 높은 文獻이 檢索된다.

LEADER에서 特記할 點은 文獻本文이 데이터베이스에 入力될 때, 自動分析되어 重要語가 가려내져서 그 文獻의 特性이 주어지며 그 文獻內에서 各 重要語가 차지하는 重要度의 加重値가 自動적으로 附與된다는 것이다. 檢索者의 檢索要求도 이와 같은 方法으로 分析되어 文獻本文 處理때에 만들어진 辭典과 對照된 다음 가장 가깝게 一致되는 單語들을 디스플레이한다. 檢索者가 이 單語들을 檢索用語로 받아들이지 않는 경우, 시스템은 이에 關連된 다른 單語群을 提供하고 이와 같은 利用者와 시스템과의 對話의 結果 利用者는 그가 願하는 文獻의 本文을 터미널에서 받아 볼 수 있게 된다. 이 시스템은 1971년 Lehigh대학의 Mart Science and Engineering Library에 LEADERMART라는 이름으로 받아들여 졌으며 하드웨어로는 CDC 6400 컴퓨터가 使用되고 있다. 1973년에는 COMPENDEX(Engineering Index)와 CHEMCON(Chemical Condensates)의 데이터베이스를 위한 LEADERMART 온라인 서어비스가 시작되었다.

#### 4.8 MEDLINE

MEDLINE(MEDLARS On-Line)은 NLM(National Library of Medicine)에 의해 始作된 시스템으로 1970년 AIM-TWX(Abridged Index Medicus by Teletypewriter Exchange Network)에서 그 起源을 찾을 수 있다. AIM-TWX는 MEDLARS가 取扱하는 1965년 以後의 약 122個 醫學分野 키어저어널에 包含된 記事에 대한 檢索시스템으로 데이터베이스는 每月更新되었다. 1972年初 약 180,000件에 다다른 이 데이터베이스는 캘리포니아州 산타모니카에 位置한 SDC(System Development Corporation)의 IBM 360/67 컴퓨터에 保管되어 TWX 혹은 電話네트워크를 通하여 利用者가 터미널에서 檢索하도록 되었다. AIM-TWX시스템에서 使用했던 檢索시스템은 事實상 ORBIT의 한 version이었다. SDC는 MEDLARS의 檢索을 위하여 ORBIT를 補強하여 ELHILL이라 하였으며 後에 이것을 온라인으로 發展시켜 ELHILL II라 하였다. 現在 쓰이고 있는 것은 ELHILL III로 NLM에 의해 使用되어 MEDLINE이라 불리우고 있다. NLM에서는 1972년 4월 이후 IBM 370/155를 使用하고 있으며 1973년 이 데이터베이스는 1970년 1월 以後의 1,200개 醫學分

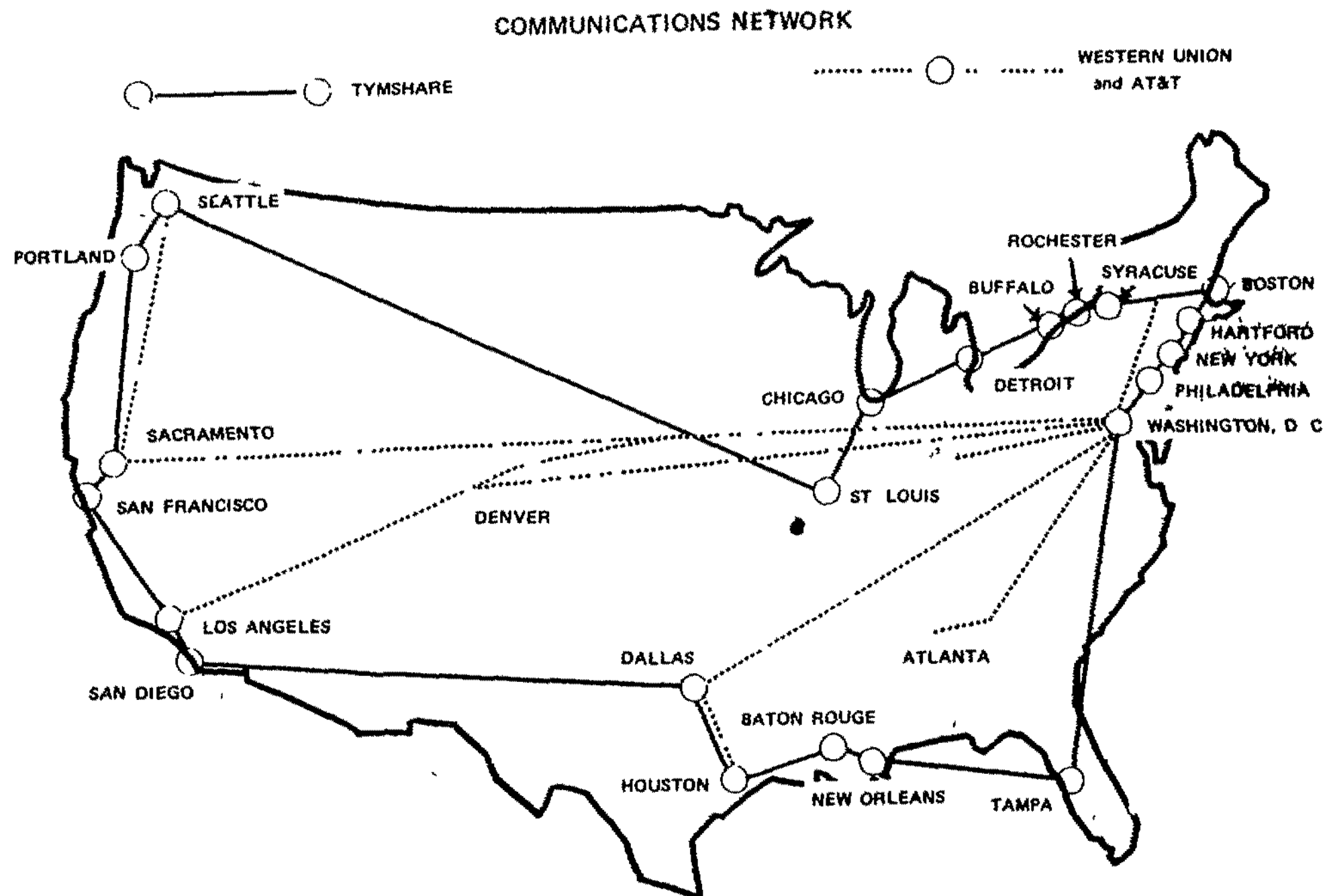


그림 4. MEDLINE 네트워크

野雜誌의 425,000件的 書誌的 事項을 包含하고 있다.

이 서비스를 받는 利用者는 터미널과 電話回線 使用料로 每月 약 \$65.00씩 支拂했으며 1973년 NLM은 MEDLINE의 使用料를 1時間 當 \$6.00로 定하였다. 커뮤니케이션에 드는 費用을 줄이기 위하여 West Union과 AT & T의 電話回線에 의해 試圖되었던 이 시스템을 1972년에는 TYMSHARE 네트워크의 利用으로 바꾸었다. TYMSHARE 네트워크의 長點은 利用者가 MEDLINE에서 使用하고 있는 電話番號를 通하여 네트워크 안의 다른 컴퓨터를 使用할 수 있다는 것이다. NLM 네트워크는 200個 機關의 약 40~45 利用者에게 同時에 서비스를 提供할 수 있다. NLM네트워크는 그림 4와 같다.

1973년에는 MEDLINE 네트워크를 利用한 데이터베이스의 다른 面의 活用이 計劃되었으며 이중의 하나가 SDILINE (Selective Dissemination of Information On-Line)이라는 것이다. SDILINE은 MEDLARS에 包含된 모든 雜誌의 最近號에 대한 記事索引 데이터베이스로 每月

發刊되는 Index Medicus에 앞서 情報를 提供한다. 此外 CATLINE은 NLM의 機械가 읽을 수 있는 形態의 目錄파일로 著者名, 逐次刊行物名, 主題名, 言語 등을 使用하여 願하는 文獻을 檢索하는 시스템이며 도서관에서 圖書購入, 目錄, 相互貸借; 參考서비스에서 有用하게 使用된다.

MEDLINE의 利用者數는 年間 약 240,000명에 달하며 그 數는 每年 增加하고 있다.

#### 4.9 New York Times Information Bank

이 시스템은 The New York Times가 300萬 \$을 들여 1966년 始作하여 1973년 完成한 프로젝트로 소프트웨어는 IBM Federal Systems Division에 의하여 만들어졌으며 使用機種 및 이에 關連된 事項은 表 5와 같다.

The New York Times에 실린 뉴스의 抄錄 및 編集事項에 關한 데이터를 온라인으로 迅速하게 찾아 보는 것을 目的으로 하는 이 시스템은 미리 選擇된 약 70개의 新聞과 雜誌에 실린 記事를 取扱한다. 데이터는 年間 약 200,000件씩 增加하는데 그중 약 100,000件이 The New York

表 5. New York Times Information Bank의 시스템 構成

Central processing unit	IBM 360/50 with a core of 524,000 bytes (which will be shared with other applications). (IBM 370/145 was under consideration in 1972.)
Mass storage devices	IBM 2314 direct access disk storage facility (200 million bytes each). IBM 2321 Data Cell Drive (400 million bytes each) (ultimately to be replaced by IBM 3330 disks).
Tape drives	IBM 2401 (model 4).
High-speed printer	IBM 1403 (uppercase and lowercase).
Microfiche storage and retrieval device	Foto-Mem RISAR (storage capacity of 4.95 million page images) controlled by a CENTAUR computer. <sup>a</sup>
In-house terminals (used for data entry, inquiry, and output)	IBM 4506 display station with 4275 terminal control unit.
Out-of-house terminals (used for inquiry and output)	Any of several compatible CRT and teletypewriter terminals.
Access method	BTAM
Operating system	DOS
Programming languages	BAL, PL/I, and COBOL.



Times로부터이다. 뉴스의 本文은 마이크로피쉬 形態로 바꾸어 컴퓨터에 連結된 마이크로필름 檢索機에 貯藏된다. The Times로부터이든 다른 雜誌로부터이든 이 데이터베이스에 실릴 資料는 온라인에 의해 抄錄과 索引이 만들어 진다. 檢索에 사용될 각 索引語에는 그 用語가 記事內에서 차지하는 重要도에 따라 加重値가 주어지며 抄錄自体도 그 記事의 重要도에 따라 加重度가 附與된다. 예를 들면 第1面の 1段記事인 경우에는 最高의 加重値인 8이 주어진다. The Times에 실린 記事는 이와 같은 方法으로 新聞 發刊後 약 72~96時間 以內에 데이터베이스에 包含되어 진다. 보통 1個의 fische에는 99個까지의 記事를 실을 수 있으며 이것은 週間 1日分の 新聞量에 該當된다. 그러나 日曜日 新聞은 1~4個의 fische를 要한다. 각 fische는 順序대로 番號가 매겨지며 新聞記事抄錄은 그 記事의 書誌事項(페이지, 欄, 記事의 形態 등)과 그 記事가 실린 fische 番號 및 그 fische內에서의 記事番號 등의 情報과 함께 컴퓨터에 保管된다. 檢索을 위하여 利用者는 그가 興味를 가진 主題를 몇개의 用語로 表現하면, 이 用語들은 시스템 內部에 保管되었던 디소오러스와 對照된 후 서로 對應되는 것이 있으면 檢索語로 받아들여지고 그렇지 않으면 그 用語에 關連된 用語를 提示하여 利用者로 하여금 그의 檢索用語를 修正하도록 한다. 따라서 冊子形態로 된 Information Bank Thesaurus는 利用者が 檢索用語를 選擇하는데 도움을 줄 수 있다. 檢索用語가 定해지면 檢索者는 이 用語들을 부울린公式으로 連結하여 檢索을 한다. 檢索者는 記事가 실린 날자, 記事形態(例: 社說, 뉴스分析), 寫眞 혹은 記事의 加重値를 사용함으로써 그의 檢索限界를 좁힐 수 있다. 檢索結果로는 書誌的 事項과 記事抄錄이 本文이 실린 microfiche의 番號와 함께 提供된다. 本文自体는 Times社內의 利用者에게는 CC(Closed Circuit) TV를 通해 CRT에 비쳐지고 願하는 경우에는 라인프린터를 通해 프린트 形態의 出力도 可能하나, 遠隔地의 利用者들에게는 이러한 設備가 不可能하다. 따라서 遠隔地의 檢索者는 microfiche reader를 사용하여 檢索된 記事의 本文을 읽을 수 있도록, The New

York Times에 실린 記事에 限하여 microfiche가 이 시스템과 함께 提供된다.

120個線에 500個까지의 터미널 連結이 可能한 이 시스템의 使用料는 通信回線 使用料와 마이크로피쉬價格을 除外하고 年間 약 \$25,000.00에 달한다.

#### 4.10 RECON

RECON(Remote Console)은 NASA(National Aeronautic and Space Administration)에서 사용하고 있는 온라인 檢索시스템으로 IBM360/50과 두개의 2314디스크 및 2개의 데이터셀이 사용되고 있다. 데이터베이스는 NASA가 發刊하는 Scientific and Technical Aerospace Reports, International Aerospace Abstracts 및 Aerospace Medicine and Biology에 실리는 모든 記事에 대한 書誌的 事項을 包含한다. 檢索者는 NASA 디소오러스에서 그의 檢索用語를 選擇한 후 부울린公式으로 이 用語들을 連結시킨다. 檢索用語는 하나씩 入力되어야 하며 시스템은 入力된 하나의 用語에 대하여 檢索될 文獻數와 디소오러스에서 그 用語에 알파벳順序로 用語들에 대한 文獻數를 提供한다. 또한 特定命令을 사용함으로써 어떤 用語에 意味上 가까운 用語들을 찾아볼 수 있도록 해 주고 있다. RECON은 NASA에 의해 Lockheed Missiles and Space Co.에서 개발된 것이기 때문에 Lockheed의 DIALOG와 거의 같다.

RECON은 情報檢索만을 위한 시스템이므로 파일維持의 機能은 갖고 있지 않다. 이 機能을 위하여 STIMS(Scientific and Technical Information Modula System)이 함께 사용되고 있다.

디소오러스에 실리지 않은 自然語(natural language)를 直接 檢索에 사용할 수 없고 檢索用語들을 처음부터 부울린公式에 의해 連結시킬 수 없는 不便을 없애기 위하여 1972년에는 RECON을 修正하여 自然語 및 用語의 斷切(truncation) 등을 사용할 수 있게 되었으며 直接 用語들을 부울린公式으로 連結시킬 수 있게 되었다. 이 시스템은 書誌情報 以外の 情報檢索에도 利

用되고 있다.

#### 4.11 State University of New York(SUNY) Bibliographic Retrieval System

SUNY Biomedical Communication Network는 최초의 Biomedical分野 文獻檢索시스템으로 1968년 9개의 醫學圖書館을 中心으로 始作되었다. SUNY/BCN은 Albany에 있는 SUNY Central Office Computer Center의 IBM 370/158을 利用하며 初期의 데이터베이스로는 1) MEDLARS 2) NLMbook(NLM의 Current目錄데이터프) 3) Netbook(SUNY 醫學도서관에 所藏된 單行本目錄)이던 것이 1973년부터 데이터베이스의 擴張과 함께 IBM의 STAIRS(Storage and Information Retrieval System)를 渡入, 使用함으로써 自然語를 통한 檢索이 可能하게 되었다. 이 중 NLMbook와 NETbook은 後에 CATLINE의 母體가 된 것으로 目錄의 共同作成을 처음으로 試圖하였다는데 그 意義가 있다. 現在 32개의 醫學도서관이 加入되어 온라인 서어비스를 받고 있는 이 시스템의 데이터베이스는 1) MEDLARS: Index Medicus, International Nursing Index, Index to Dental Literature에 包含된 세계 2,700種의 雜誌로부터의 약 2백만件的 記事에 관한 情報로 1964년까지 溯及調査(Retrospective Search)가 可能하다. 이 SUNY/MEDLARS 데이터베이스에 대한 檢索은 MeSH (Medical Subject Heading)에 收錄된 用語를 檢索用語로 使用해야 하며 雜誌名, 著者名 등에 의해 그의 檢索을 制限할 수 있다. 2) ERIC: 教育關係의 데이터베이스로 RIE(Research in Education,) CIJE(Current Index to Journals in Education)에 收錄된 記事에 대한 情報를 提供한다.

RIE파일은 1966년 以後의 70,000件的 記事를 包含하며 每年 약 12,000件씩 그 數가 增加하고 있다. CIJE파일은 1969년 以後의 약 90,000件的 記事를 取扱하며 每年 약 18,000件씩 增加하므로 ERIC에는 總 160,000件的 記事가 包含

된다 하겠다. SUNY/ERIC은 Thesaurus of ERIC Descriptor에 收錄된 主題語를 使用하거나, 自然語形態의 檢索要求를 抄錄本文, 記事名, 著者名과 對照함으로써 探索할 수 있다.

3) Psychological abstracts: 1967년 以後 Psychological abstracts에 收錄된 160,000件的 記事에 대한 情報로 800種의 心理學, 行動科學分野 雜誌, 單行本, 研究報告書 등 行動 및 社 社科學의 情報를 取扱한다. SUNY/PSYCH ABSTRACTS 데이터베이스는 Thesaurus of Psychological Terms에 收錄된 主題語를 使用하여 檢索하거나 自然語形態의 檢索要求를 抄錄本文, 記事名, 著者名 등과 對照하여 檢索을 行할 수도 있다.

#### 參 考 文 獻

- 1) F. W. Lancaster and E. G. Fayan, Information Retrieval On-Line, Los Angeles, Melville Publishing, 1973.
- 2) Charles T. Meadow, The Analysis of Information Systems, Los Angeles, Melville Publishing, 1973.
- 3) C. F. J. Overhage and J. F. Reintjes, Project Intrex: a general review., Information Storage and Retrieval Vol. 10, No. 516, 1974. pp. 157~188.
- 4) Egeland, Janet, The SUNY Biomedical Communication Network: Six Years of Progress in On-Line Bibliographic Retrieval., Bulletin of the Medical Library Association, Vol. 63, No. 2, 1975. pp. 189~194
- 5) Wood, M. Sandra, and Seeds, Robert S., Development of SDI Services from a Manual Current Awareness Service to SDILINE., Bulletin of the Medical Library Association, Vol. 62, No. 4, 1974. pp. 374~384
- 6) Susanne M. Humphrey, Searching the MEDLARS citation file on-line using ELHILL and STAIRS: an updated comparison, Information Processing and Management, Vol. 12, No. 1, 1976. pp. 63~70