

# TECHNOLINE 情報檢索시스템 STAIRS 紹介

李 惠 成

〈KORSTIC 電算室〉

## 1. 序 論

20世紀 후반부터 知識의 發達과 科學技術의 急進的인 發展으로 인하여 情報의 量이 急激히 늘어나게 되었다. 이러한 情報量의 增加와 더불어 情報의 要求 또한 顯著히 많아졌으며 따라서 有用한 各種 情報를 適切히 加工, 蓄積하여 언제 든지 情報를 必要로 하는 利用者에게 提供할 수 있는 体制가 必要하게 되었으며 컴퓨터의 急速한 發展은 컴퓨터에 의한 情報檢索手段을 提供하게 되었다.

韓國科學技術情報센터(KORSTIC)는 지난 75 年末부터 KIST의 컴퓨터를 使用하여 약 500餘 利用者에게 컴퓨터에 의한 情報檢索서비스를 實施하여 왔다. 78年 11月에는 IBM370/138(512KB) 을 設置稼動하여 CAC(Chemical Abstract Condensates), INSPEC(Information Services for Physics, Electrotechnology, Computer and Control), ISMEC(Information Services for Mechanical Engineering)에 이어 COMPENDEX (Computerized Engineering Index) 등 數種의 데이터베이스에 대해 종전의 SDI(Selective Dissemination of Information)서비스 뿐만 아니라 온라인에 의한 RS(Retrospective Search)를 實施하게 되었다. 또한 이를 發展시킨 데이터뱅크計劃을 樹立 온라인情報檢索서비스를 TECHNOLINE이라 命名하였다.

· 컴퓨터에 의한 情報檢索 프로그램으로는 STAIRS(Storage and Information Retrieval System), FAIRS(FACOM Advanced Information Retrieval System), DIR(Document Information Retrieval), IRMS(Information Retrieval and Management System) 등과 같이 既成商品화된 檢索시스템과 Lockheed의 DIALOG, SDC의 ORBIT 등 情報서비스機關에서 스스로 開發하여 使用하고 있는 두가지로 나눌 수 있다.

現在 KORSTIC에서 使用하는 檢索시스템은 컴퓨터 製造會社인 IBM社에서 開發한 STAIRS로서 KORSTIC이 國內 最初로 導入하였다. 本稿에서는 STAIRS의 시스템構造 및 性能에 대하여 叙述하고자 한다.

## 2. STAIRS시스템構成

STAIRS는 SEARCH, SELECT, RANK 등의 다양한 探索機能을 가지고 있으며 배치방식이나 온라인방식으로 檢索할 수 있다. 특히 온라인 檢索으로 遠距離에서도 터미널을 이용하여 檢索者가 中央컴퓨터의 방대한 情報를 對話式으로 檢索할 수 있다. 既存 데이터베이스에 새로운 데이터베이스를 쉽게 追加한다든지 데이터베이스 維持 및 管理를 위한 여러가지 배치 프로그램을 包含하고 있으며 STAIRS 시스템 構成圖는 그림 1과 같다.

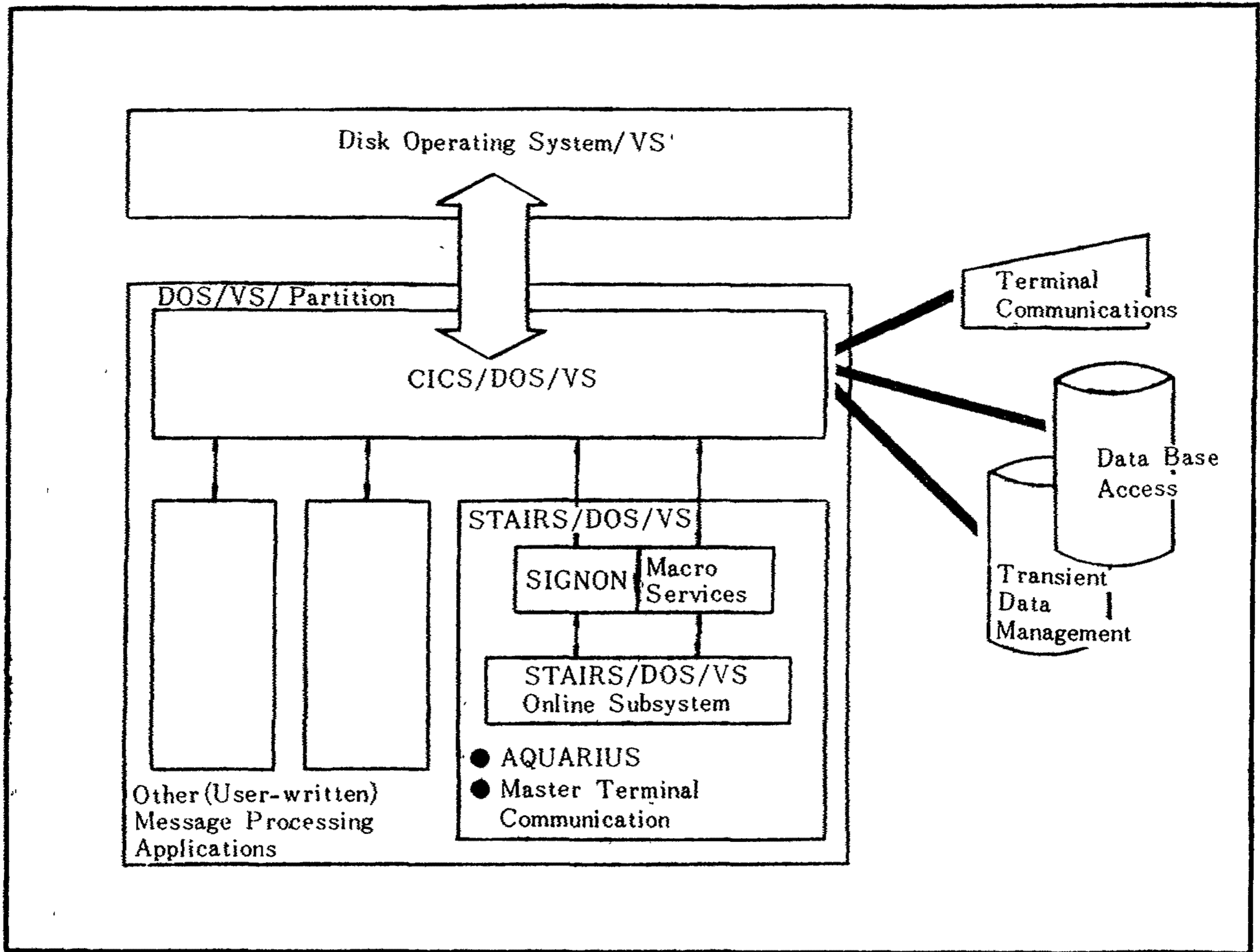


그림 1. STAIRS 構成圖

### 3. STAIRS 데이터베이스 構造

STAIRS에서의 데이터베이스 레코드의 構成 概念은 WORD, SENTENCE, PARAGRAPH 와 FORMATTED FIELD로 이루어지며 一般 的인 Document의 構成原則을 그대로 維持하고 있으므로 文脈探索(Text Search)이 可能하다. STAIRS 데이터베이스는 크게 나누어 다음과 같은 다섯가지 파일들로 이루어지며 이들의 內容 및 構造는 그림 2와 같다.

- 本文파일(Text File) ...프린트되거나 디스플레이되는 원래의 도큐먼트와 거의 같은 形態이며 레코드의 길이는 1966Byte이다.

- 本文索引파일(Text Index File) ...Text 파일에 대한 Index파일로서 FORMATTED FIE-

LD의 內容을 包含하고 있으며 검색명령중 SELECT에 주로 사용되는 파일이며 레코드의 길이는 72Byte이다.

- 辭典體파일(Dictionary File) ...데이터베이스內的 모든 單語들의 Inverted파일에 대한 Pointer와 데이터베이스內에서의 그 單語의 出現頻度數를 包含하고 있으며 레코드의 길이는 1964 Byte이다.

- 變換파일(Inverted File) ...各 單語들의 데이터베이스內에서의 出現頻度數와 各 Document에서의 그 單語의 위치를 나타내고 있는 파일로서 레코드의 길이는 1960Byte이다.

- 기타 補助파일 ...各 데이터베이스에 대한 Description파일, Message파일, 利用者登錄파일 등이 있다.

데이터베이스를 分析하여 위와 같은 STAIRS

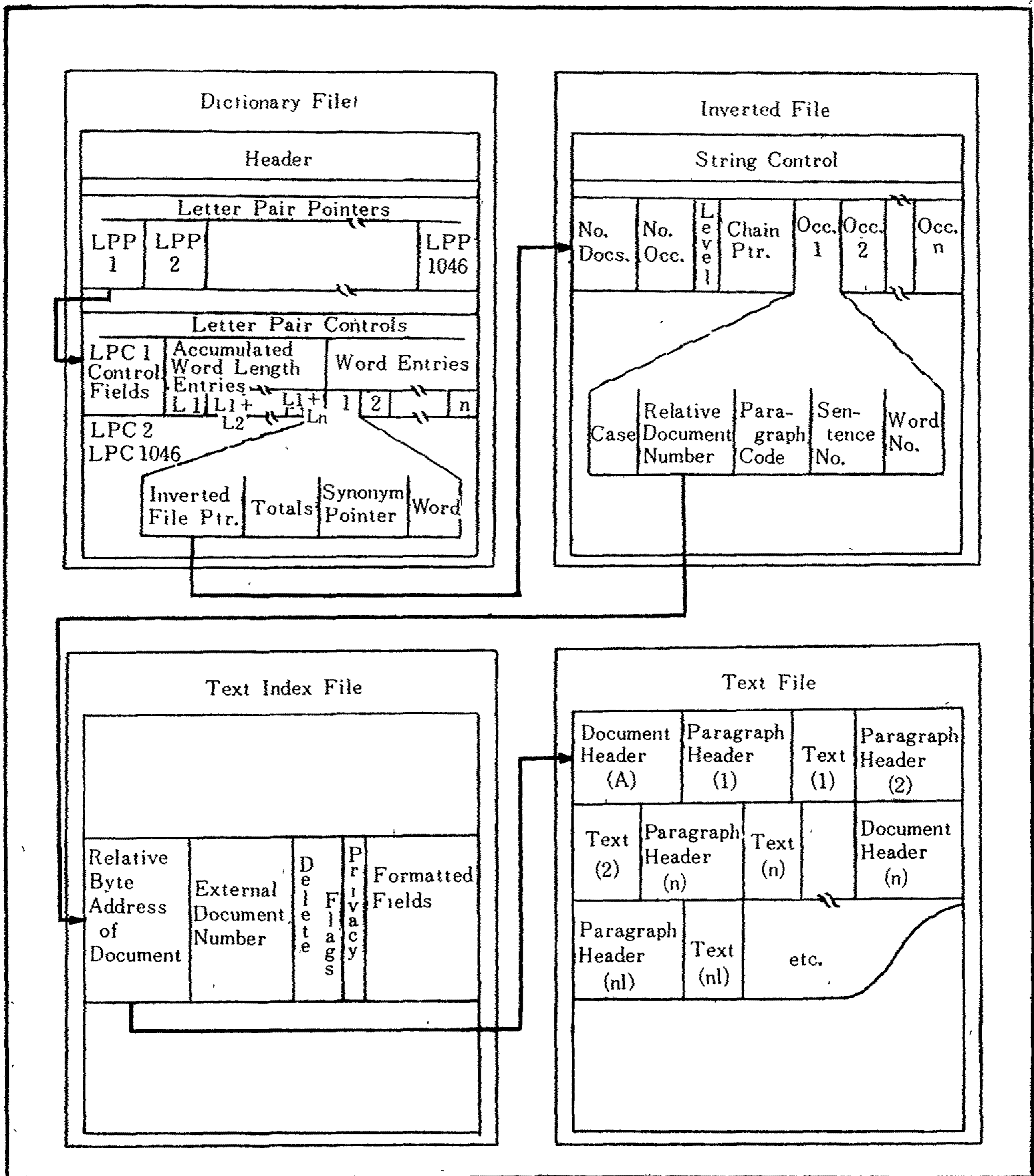


그림 2. STAIRS 데이터베이스 구조

파일들을 만들기 위한 배치 프로그램들도 STAIRS에 포함되어 있으며 이들 파일作成은 檢索作業과는 별도로 배치作業으로 행해져야 한다. 데이터베이스化 하고자 하는 情報 또는 데이터베이스들은 特定 Format(EDIO, CNDTXT)을

가지고 있어야 하며 KORSTIC에서는 入手 데이터베이스를 일단 CNDTXT(Condensed Text)形態로 바꾼 다음 STAIRS에 의해 STAIRS 데이터베이스를 作成한다. 데이터베이스作成 Flow는 그림 3과 같다.

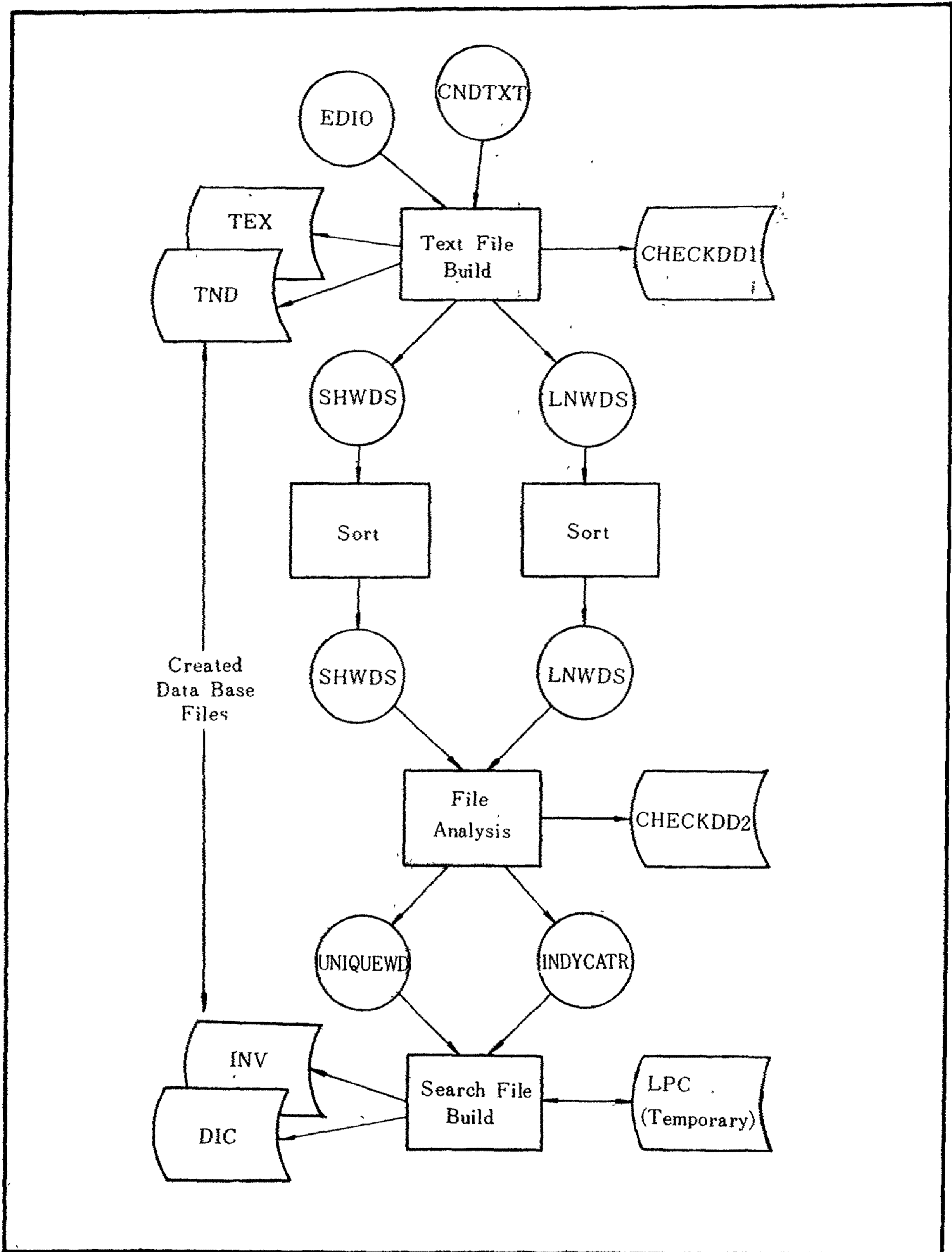


그림 3. 데이터베이스作成 Flow

AQUA PSWD USER NAME

R0102 ENTER DATA BASE NAME INSP

R0201 ENTER ONE OF THE FOLLOWING COMMANDS:

..BROWSE ..SEARCH ..SELECT ..EXEC ..RANK ..HELP ..CHANGE ..OFF  
 ..SEARCH

AQUARIUS - SEARCH MODE - BEGIN YOUR QUERY AFTER THE STATEMENT NUMBER

00001 QUALITY AND CONTROL OR QC ----- ①  
 QUALITY 383 OCCURRENCES 209 DOCUMENTS  
 CONTROL 2674 OCCURRENCES 1002 DOCUMENTS  
 QC 0 OCCURRENCES 0 DOCUMENTS  
 RESULT 62 OCCURRENCES 62 DOCUMENTS

AQUARIUS - SEARCH MODE - BEGIN YOUR QUERY AFTER THE STATEMENT NUMBER

00002 QUALITY SAME CONTROL ----- ②  
 QUALITY 383 OCCURRENCES 209 DOCUMENTS  
 CONTROL 2674 OCCURRENCES 1002 DOCUMENTS  
 RESULT 85 OCCURRENCES 47 DOCUMENTS

AQUARIUS - SEARCH MODE - BEGIN YOUR QUERY AFTER THE STATEMENT NUMBER

00003 QUALITY WITH CONTROL ----- ③  
 QUALITY 383 OCCURRENCES 209 DOCUMENTS  
 CONTROL 2674 OCCURRENCES 1002 DOCUMENTS  
 RESULT 86 OCCURRENCES 47 DOCUMENTS

AQUARIUS - SEARCH MODE - BEGIN YOUR QUERY AFTER THE STATEMENT NUMBER

00004 QUALITY ADJ CONTROL ----- ④  
 QUALITY 383 OCCURRENCES 209 DOCUMENTS  
 CONTROL 2674 OCCURRENCES 1002 DOCUMENTS  
 RESULT 75 OCCURRENCES 37 DOCUMENTS

AQUARIUS - SEARCH MODE - BEGIN YOUR QUERY AFTER THE STATEMENT NUMBER

00005 QUALITY.TITLE, ----- ⑤  
 QUALITY 383 OCCURRENCES 209 DOCUMENTS  
 RESULT 40 OCCURRENCES 39 DOCUMENTS

AQUARIUS - SEARCH MODE - BEGIN YOUR QUERY AFTER THE STATEMENT NUMBER

00006 ABSTRACT\$ ----- ⑥  
 ABSTRACT\$ 251 OCCURRENCES  
 ABSTRACT 224 OCCURRENCES 214 DOCUMENTS  
 ABSTRACTED 1 OCCURRENCE 1 DOCUMENT  
 ABSTRACTING 1 OCCURRENCE 1 DOCUMENT  
 ABSTRACTION 10 OCCURRENCES 5 DOCUMENTS  
 ABSTRACTLY 1 OCCURRENCE 1 DOCUMENT  
 ABSTRACTS 14 OCCURRENCES 10 DOCUMENTS  
 RESULT 251 OCCURRENCES 230 DOCUMENTS

AQUARIUS - SELECT MODE - ENTER EXTENT AND CRITERIA AFTER THE STATEMENT  
 NUMBER OR ENYLR EXPLAIN ----- ⑦

00007 3 ABST.NO WL A78000000,A78100000  
 J2516 NUMBER OF SELECTED DOCUMENTS = 0000/4

..OFF

그림 4. 檢索例



AQUARIUS - BROWSE MODE - ENTER PRINT SPECIFICATIONS - OR:  
ALL - TOTAL DOCUMENT, A - ALL PARAGRAPHS, B - ALL FIELDS, OR C OR EXPLAIN  
ALL

```

INSP01264944 DOCUMENT=      11 OF      74      PAGE =      1 OF      1
  ABST.NO   = A78090254
  LANGUAGE  = ENGLISH
  PUB.TYPE  = C
  VOL.ISS   = 7823
TITLE      PROCEEDINGS OF THE INTERNATIONAL CONFERENCE AND SYMPOSIUM ON
           ELECTRON TUNNELING.
CITATION    (DATE: 25-27 MAY 1977   DATE: 1978   SBN: J 540 08691 9
SECNO: A3390; A0130C; A3580   ASTNO: A78090254   TR: G.
SUBJECT     TUNNELLING SPECTROSCOPY.
KEYWORD     IETS; ELASTIC TUNNELLING; INELASTIC ELECTRON TUNNELLING
           SPECTROSCOPY.
ABSTRACT    THE FOLLOWING TOPICS WERE DEALT WITH: APPLICATIONS OF IETS:
           THEORETICAL ASPECTS OF ELECTRON TUNNELLING; MOLECULAR ADSORPTION ON
           NONMETALLIC SURFACES; NEW APPLICATIONS OF IETS; AND ELASTIC
           TUNNELLING. 27 PAPERS WERE PRESENTED, OF WHICH 24 ARE PUBLISHED IN
           FULL IN THE PRESENT PROCEEDINGS, AND 3 AS ABSTRACTS ONLY.
           END OF DOCUMENT
    
```

```

INSP01265722 DOCUMENT=      17 OF      74      PAGE =      1 OF      1
  ABST.NO   = A78091222
  LANGUAGE  = ENGLISH
  PUB.TYPE  = C
  VOL.ISS   = 7823
TITLE      PHYSICS AND APPLICATIONS OF SMECTIC AND LYOTROPIC LIQUID
           CRYSTALS.
CITATION    (DATE: 9-13 JAN. 1978   DATE: 1978   SECNO: A6130; A6470L;
A0130C   ASTNO: A78091222   TR: G   CODEN: ANPHA.
SUBJECT     SMECTIC LIQUID CRYSTALS; LIQUID CRYSTALS; LIQUID CRYSTAL PHASE
           TRANSFORMATIONS.
KEYWORD     SMECTIC LIQUID CRYSTALS; LYOTROPIC LIQUID CRYSTALS;
           PLASTICITY; ELECTRIC FIELD EFFECTS; LIQUID CRYSTAL DEFECTS, TEXTURAL
           INSTABILITIES; LIQUID CRYSTAL PHASE TRANSITIONS.
ABSTRACT    THE FOLLOWING TOPICS WERE DEALT WITH: SMECTIC AND LYOTROPIC
           MATERIALS, STRUCTURE AND DYNAMIC PROPERTIES OF SMECTICS, PLASTICITY
           AND DEFECTS, ELECTRIC FIELD EFFECTS, LYOTROPIC SYSTEMS, TEXTURAL
           INSTABILITIES AND APPLICATIONS, AND PHASE TRANSITIONS. 56 PAPERS
           WERE PRESENTED, OF WHICH 34 ARE PUBLISHED IN FULL IN THE PRESENT
           PROCEEDINGS, AND 22 AS ABSTRACTS ONLY.
           END OF DOCUMENT
    
```

그림 5. STAIRS 檢索出力

#### 4. 檢索機能

STAIRS 를 이용하여 情報를 檢索하고자 하는 利用者는 利用者登錄 파일에 그의 Password, 利用者 코드, 데이터베이스名이 등록되어 있어야 한다. 檢索機能에는 크게 나누어 네가지 類型이 있으며 類型別 命令語 種類는 다음과 같다.

- 探索...SEARCH, SELECT, RANK
- 디스플레이...DISPLAY, BROWSE
- 프린트...PRINT, COPY, MAIL
- 기타...SAVE, EXEC, CHANGE, HELP 등

각 命令語의 內容은 表 1 과 같으며 檢索機能 중 SEARCH 및 SELECT에 대해서 자세히 說明하겠다.

表 1. 命令語 種類와 内容

- ..SEARCH : 探索
- ..SELECT : Formatted Field값에 의한 探索制限
- ..BROWSE : 探索된 情報를 디스플레이
- ..RANK : 加重値에 의한 探索
- ..CHANGE : 데이터베이스 變更
- ..PRINT : OFF-LINE PRINTER로 檢索된 情報를 프린트
- ..HELP : STAIRS 檢索機能에 대한 說明
- ..SORT : Formatted Field의 内容에 의한 Sort
- ..MAIL : OFF-LINE PRINTER로 檢索된 情報를 傳達
- ..EXEC : SAVE되어 있는 質問을 遂行
- ..COPY : 터미널 프린터로 프린트
- ..SAVE : 檢索質問을 나중에 再使用하기 위하여 保管

..SEARCH... 一般的인 探索機能으로 檢索用語를 入力시키면 데이터베이스內에서 探索하고자 하는 檢索用語의 出現頻度數와 檢索된 文獻의 件數를 알려준다. 그림 4-①과 같이 檢索質問의 各 用語사이를 檢索論理(AND, OR, NOT)의 使用으로 이어나갈 수 있으며 그림 4-②에서와 같이 SAME을 使用하여 QUALITY와 CONTROL 두 單語가 Paragraph 單位內에 있을 때 檢索해 낸다. 그림 4-③에서의 WITH를 使用하여 QUALITY와 CONTROL이 Sentence 單位內에 있을 때 檢索해 내며 그림 4-④와 같이 ADJ(adjacent)를 使用하여 QUALITY와 CONTROL이 바로 인접해 있는 경우라도 檢索할 수 있다. 그림 4-⑤와 같이 하면 TITLE Paragraph內에 QUALITY가 있는 경우만 檢索이 된다. 그림 4-⑥과 같이 "\$"를 使用하면 ABSTRACT로 시작되는 辭典體파일內의 모든 單語에 대해 檢索한다.

..SELECT... 檢索命令中 SEARCH는 檢索用語가 一致되는 것만을 檢索해 내는 반면에 SELECT는 값의 比較가 可能하다. SELECT는 Formatted Field에 定義되어 있는 Field에만 可能하며 特殊한 演算子(LT...Less Than, EQ...Equal to, GT...Greater Than, WL...Within the Limits) 등을 使用한다. 그림 4-⑦에서와

같이 윗 문장인 4-⑥에서 SEARCH를 使用하여 檢索한 結果中 ABST.NO라고 定義된 Formatted Field의 값이 WL 즉 A78000000와 A78100000 사이에 包含되는 경우만 檢索된다.

### 5. 소프트웨어 및 하드웨어

STAIRS는 基本的으로 온라인 檢索用이기 때문에 CICS(Customer Information Control System)와 같은 온라인 통제시스템(On-line Control System)이 必要하다. 프로그램들은 Macro, Online, Batch 세 그룹으로 나누어지며 Macro와 Online은 Assembler Language로, Batch는 PL/1 Language로 되어 있다. 데이터베이스 作成을 위한 入力形態에 맞는 파일을 만들기 위하여 各 데이터베이스別 形態變換프로그램(Format Conversion Program)이 必要하다.

CICS-STAIRS 運營을 위하여 컴퓨터의 Main Memory가 최소한 240K Byte정도가 필요하며 최소주변장치 및 데이터베이스別 디스크용량은 表 2와 같다.

- CARD READER & LINE PRINTER
- DISK UNIT
- MAGNETIC TAPE UNIT
- TERMINAL (DISPLAY, OFF-LINE PRINTER)

表 2. 데이터베이스別 DISK 容量

데이터 베이스	레코드數 (月 평균)	레코드當 평균길이	용량 (Byte)
CAC	37,400	800(Byte)	45M
INSPEC	12,000	1300 "	30M
ISMEC	1,250	800 "	1.5M
COMPENDEX	8,000	1500 "	20M

### 6. STAIRS의 特徵, 問題点 및 發展計劃

檢索시스템에는 디소오러스나 索引語에 의해서만 檢索할 수 있는 것과 自然語 그대로 檢索할 수 있는 두가지 形態가 있다. IRMS인 경우

索引語파일은 別途로 作成 維持되어야 하며 索引語로 定義되어 있는 用語에 한하여만 檢索할 수 있다. STAIRS의 特徵은 索引語파일과 類似한 辭典體파일에 의해 檢索한다. 이 辭典體파일은 데이터베이스作成 過程에서 檢索에 不必要한 前置詞, 接續詞, 冠詞들과 같이 STOPWORD로 定義된 單語를 除外한 모든 單語에 대해 자동적으로 만들어지며 特定 Paragraph도 STOP PARAGRAPH로 定義해 줄 수 있어 探索時間을 節約할 수 있다. FORMATTED FIELD의 內容에 대해서는 단순히 一致에 의한 檢索이 아니라 GE, LT, WL, EQ 등의 演算子를 使用하여 값의 比較를 통한 檢索을 할 수 있다. STAIRS에는 데이터베이스 更新, 利用者管理, 使用時間 등 각종 統計資料들을 提供하는 프로그램을 包含하고 있다.

STAIRS의 問題點으로는 PASSWORD의 管理가 어렵고 다른 어떤 시스템에서는 用語에 대

한 EXPAND機能이 있어 檢索하고자 하는 用語 외에도 그 用語에 대한 上位, 下位 혹은 關聯語들에 대해 알아볼 수 있는 시스템이 있으나 現在 STAIRS에서는 단순히 同義語들만을 定義해 줄 수 있을 뿐이다. 또한 SELECT 檢索은 Sequential Search 方式으로 探索을 하기 때문에 應答時間이 상당히 길다.

TECHNOLINE의 앞으로 發展計劃은 81년도 까지 서울, 부산, 대구, 광주 네 곳에 스위칭센터(Switching Center)와 전국 주요도시에 21개의 터미널을 설치하여 전국적인 기술정보유통망 체제를 확립하는 것이다. 그러기 위해서는 現在의 LOCAL TERMINAL에서 REMOTE TERMINAL로의 轉換 및 TERMINAL CONTROL SYSTEM에 대한 많은 研究가 있어야 함은 물론 디스크 容量을 擴張하여 可能한 많은 情報를 同時に 處理할 수 있도록 하여야 할 것이다.