인구·보건·사회 문헌정보 검색시스템개발

송 태 민*

한국보건사회연구원은 인구·보건·사회복지분야의 전문연구기관으로서 당원 도서실에 소장하고 있는 관련분야의 전문도서, 연구보고서, 회의보고서, 학위논문, 정부간행물, 학술잡지논문의 내용을 분석하고 축적하여 연구자 및 관련분야 전문가에게 보다 신속하고 적합한 보건사회 문헌정보를 제공할 수 있는 문헌정보 검색시스템을 개발하였다. 본 논문에서는 현재 운영되고 있는 정보 검색시스템의 설계와 평가에 대해서 논의 하였다.

I. 서 론

인구·보건·사회분야의 정보량이 증가함에 따라 과거와는 달리 정보를 요구하는 이용자의 정보요구가 다양하게 되었다.

그러나 이러한 정보를 수집, 관리하고 있더라도 사용자가 쉽게 접근하여 원하는 정보를 원하는 형태로 신속하게 제공받을 수 없다면 그 정보는 효용가치가 없어지게 된다. 따라서 관련된 정보자료를 일관성있고 체계적으로 조직관리하는 정보관리업무의 연구개발이 절실히 요구된다. 현재 국내에서는 각분야별로 연구기관, 단체, 대학등에서 보다 적극적인 정보관리 및 서비스를 위하여 정보자료 관리업무의 전산화를 활발히 수행하고 있다.

한국보건사회연구원에서는 1985년 부터 "인구 및 보건의료 데이타 베이스 운용체계개발 연구"의 일환으로 관련 문헌정보를 수집, 정리, 분석하여 보건사회정보를 효율적으로 이용할 수 있는 문헌정보 검색시스템을 개발하여 동 분야 연구자에게 ON-LINE 검색을 통하여 보건·사회 문헌에 관한 정보를 제공해 주고 있다. 현재까지 약 15,000건의 전문도서, 연구보고서, 회의보고서, 학위논문, 정부간행물, 학술잡지논문이 본 시스템에 저장되어 있으며 매년 약 2,500건의 문헌이 추가로 저장된다.

따라서 본고는 검색시간이 많이 소요된 초기의 문헌정보 검색시스템을 보완하고 검색종류 및 검색기법을 향상시킴으로써 연구자에게 다양한 검색을 통해 보다 신속하고 적절한 문헌정보를 제공할 수 있는 체계적인 검색시스템을 구축하는데 그 목적이 있다.

* 本院 責任電算員

Ⅱ. 시스템 개발

본 시스템을 개발 하는데는 시스템을 단계적으로 세분화하여 시스템분석, 시스템설계, 시스템평가의 과정을 거쳐 하나의 시스템을 완성하는 시스템적 접근 방법을 사용하였으며, 검색시간이 많이 소요된 초기의 시스템을 보완하기 위하여 클러스터파일 검색과 같은 새로운 검색기법을 도입하여 응답시간(turn around time)을 줄이는데 역점을 두었다.

본 시스템에 사용한 컴퓨터는 MV 15000-8 COMPUTER SYSTEM이며 프로그램언어로는 프로그램작성이 간편하고 다중파일의 접근(multi-file access)이 비교적 쉬운 FORTRAN 77을 사용하였다. 또한 시스템의 개발과 유지보수를 쉽게 하기 위해 도식화된 도구로써 플로우차트(flow-chart)와 HIPO(Hieraechy Plus Input-Process-Otput : 시스템의 기능을 계층적으로 도식화하는 것으로 시스템모델을 INPUT, PROCESS OUTPUT으로 나타낸다)를 사용하였다.

Ⅲ. 시스템 분석

검색시스템에 입력되는 자료는 당원 도서실에서 소장하고 있는 23,380권의 전문도서(단행본, 연구보고서, 학위논문)와 391종의 연속간행물(전문잡지, 학회지, 대학논문집) 및 6,000부의 비도서자료(소책자, 회의자료, 학술지논문복사물 등)중 한국의 인구·보건·사회복지분야에 관련된 문헌을 선정한것으로 주제분석요원은 선정한 자료의 내용을 분석하고 통제어휘집(controlled vocabulary)을 사용하여 색인어를 추출한 후 정보입력용지에 자료입력 내용을 작성한다. 정보입력용지에 기록된 문헌의 정보는 전산실의 자료입력 요원을 통하여 컴퓨터에 입력된 후 입력마스터파일(input master file)에 차례로 저장된다. 자료의 입력이 모두 완료된 후 시스템 관리자는 통제어휘집을 수록한 사전파일(dictionary file)과 작업용 파일인 트랜잭션파일(transaction file)을 사용하여 자료의 오류를 수정하고 검색에 필요한 도치파일 (inverted file)과 색일파일(index file) 및 검색마스터파일(retrieval master file)을 생성하여 검색시스템에 저장한다. 검색시스템에 저장된 문헌 정보에 관한 요구가 있을 때는 시스템사용자는 요구정보를 분석한 후 검색키를 작성하고 해당 검색키를 검색시스템에 입력함으로써 검색결과를 얻을 수 있다.

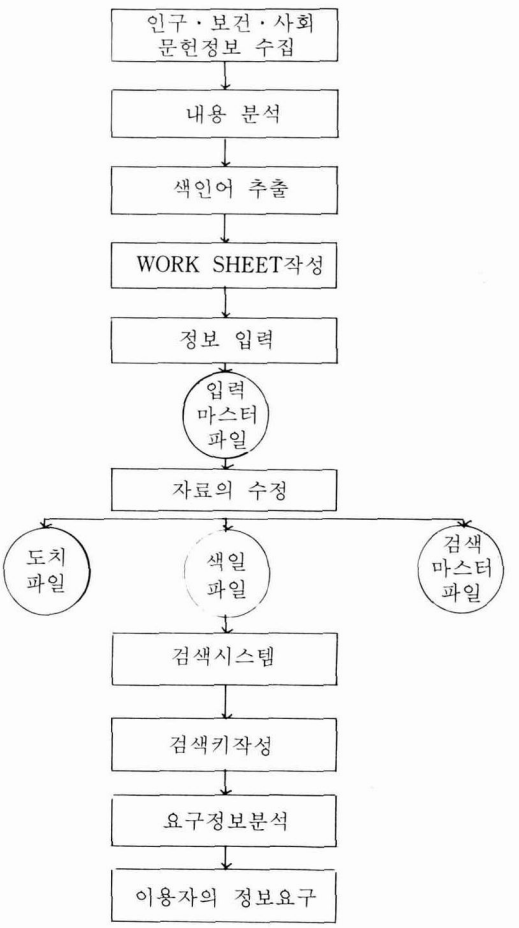

Fig. 1. Task Flow-chart of Retrieval System
검색시스템의 업무흐름도

Ⅳ. 시스템 설계

1. 입력설계(input design)

1) 입력자료항목

입력항목의 대부분은 한국목록규칙(KCR : Korean Cataloguing Rules)을 준용하여 선정하였으며 입력항목의 내용은 〈표 1〉과 같다.

2) 자료의 입력

앞에서 언급한 각 항목은 입력프로그램을 이용하여 차례로 컴퓨터에 입력되며 입력자가 쉽고 신속하게 입력할 수 있도록 〈그림 2〉와 같은 메뉴화면을 통하여 입력 및 수정을 할 수 있게 만들었다. 입력프로그램이 수행되면 입력화면의 첫번째 필드(field)에 커서(cursor)가 위치하고 SEQ-NO는 1회 입력후 자동적으로 계산되어 입력된다. 색인어(descriptor)까지 모든 입력이 완료되면 커서는 FUN 필드에 위치하게 된다. 각 입력항목에 오류(error)가 없을 경우에는 ENTER키를 입력하여 자료를 저장하고 오류가 발생했을 경우에는 오류가 있는 항목의 번호를 입력하여 그 위치에서 오류를 수정한 후 다시 FUN필드로 되돌아와 자료를 저장한다.

3) HIPO

입력, 처리과정, 출력을 그림으로 표시하면

Table 1. Input Data Items
입력자료항목

Field (Korean)	Field (English)	Length (Byte)	Content
일련번호	Citation Number	I5	
청구번호	Call Number	A14	분류번호와 저자기호로 되어 있으며 검색된 문헌의 서가위치를 알려줌
주제코드	Subject Code	A8	한문헌당 2개의 주제코드를 입력
문헌형태	Publication Type	A1	J : 잡지논문 B : 단행본 R : 연구보고서 T : 학위논문 C : 회의보고서 G : 정부간행물
저자	Author	A20 * 3	공동저자일 경우 3명까지 입력
제목 I	Title I	A150	
제목 II	Title II	A150	제목이 길거나 부제가 있을 경우 입력
출판지	Place	A4	
출판사	Publisher	A35	
연도	Year	A4	
면수	Page	A7	
언어	Language	A2	KO : 한국어 KE : 한국어 영어초록 EN : 영어 EK : 영어 한국어초록
색인어	Descriptor	A30 * 10	최대한 10개의 색인어를 입력하며 인구 및 보건분야는 미국 국립의학도서관 발행의 의학주제명표, 사회분야는 한국보건사회연구원에서 발행한 사회복지 용어집을 사용하여 추출함

THE INPUT PROGRAM OF POPULATION AND HEALTH RETRIEVAL SYSTEM

1. SEQ－NO :　　2. C－NO :　　3. SCODE :　　4. PTYPE :
5. AUTHOR1 :
6. AUTHOR2 :
7. AUTHOR3 :
8. TITLE1 :
9. TITLE2 :
P. PLACE :　　Q. PLBLISHER :
Y. YEAR :　　A. PAGE :　　L. LANGEAGE :
D. DESCRIPTOR PART
1. KEY1 :　　2. KEY2 :
3. KEY3 :　　4. KEY4 :
5. KEY5 :　　6. KEY6 :
7. KEY7 :　　8. KEY8 :
9. KEY9 :　　10. KEY10 :

FUN(EDIT : P, Q, Y, A, L, D) :

Fig. 2. Input Screen of Bibliographic Data
문헌자료의 입력화면

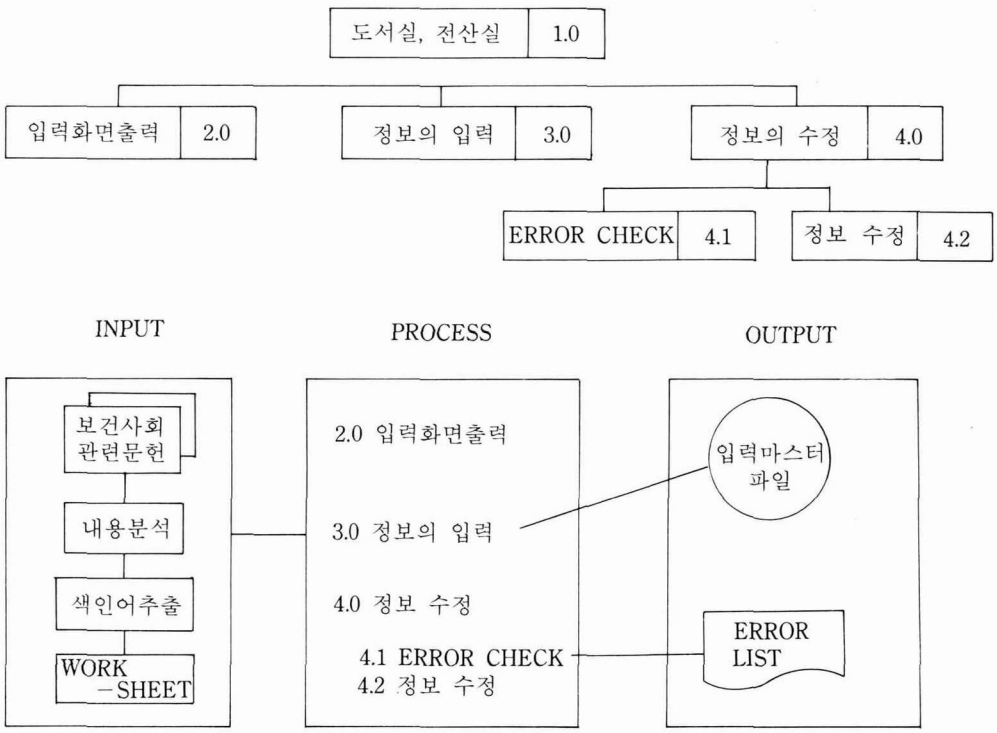

Fig. 3. HIPO of Input Design
입력설계의 HIPO

〈그림 3〉과 같다.

2. 출력설계(output design)

1) 출력설계

출력자료는 CRT의 화면 및 하드카피프린터(hard copy printer)로 출력되며 CRT화면에는 2건씩, 하드카피프린터에는 출력라인을 조정하여 A4용지를 기준으로 7~10건씩 출력하도록 설계되어있다. 검색된 문헌은 〈그림 4〉와 같이 잡지논문일 경우에는 A : 일련번호, B : 저자, C : 문헌형태, D : 서명 E : 소재지명, F : 통권수, G : 수록페이지, H : 출판지, I : 출판사, J : 출판년도, K : 언어, L : 색인어 순으로 출력되고 일반도서는 A : 일련번호, B : 저자, C : 문헌형태, D : 청구번호, E : 서명, F : 출판지, G : 출판사, H : 출판년도, I : 면수, J : 언어, K : 색인어 순으로 출력된다.

2) HIPO

입력, 처리과정, 출력을 그림으로 표시하면 〈그림 5〉와 같다.

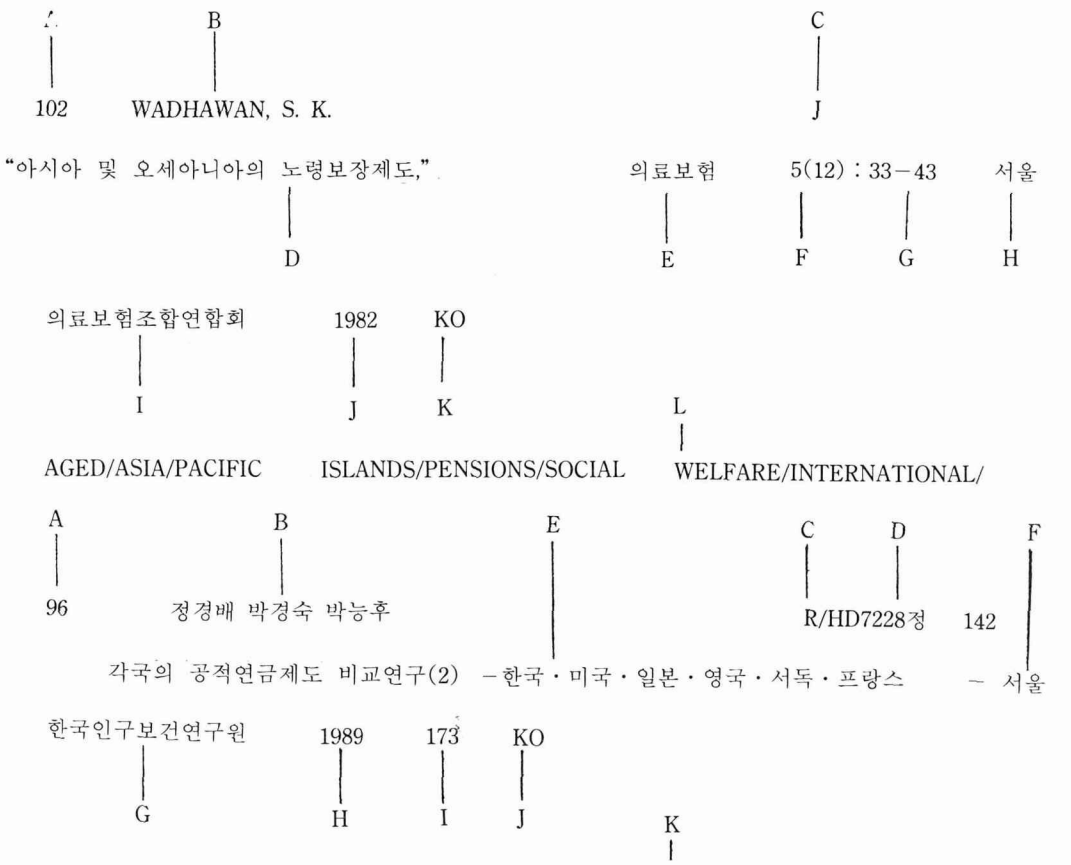

Fig. 4. Output Form of Bibliographic Data
문헌자료의 출력형태

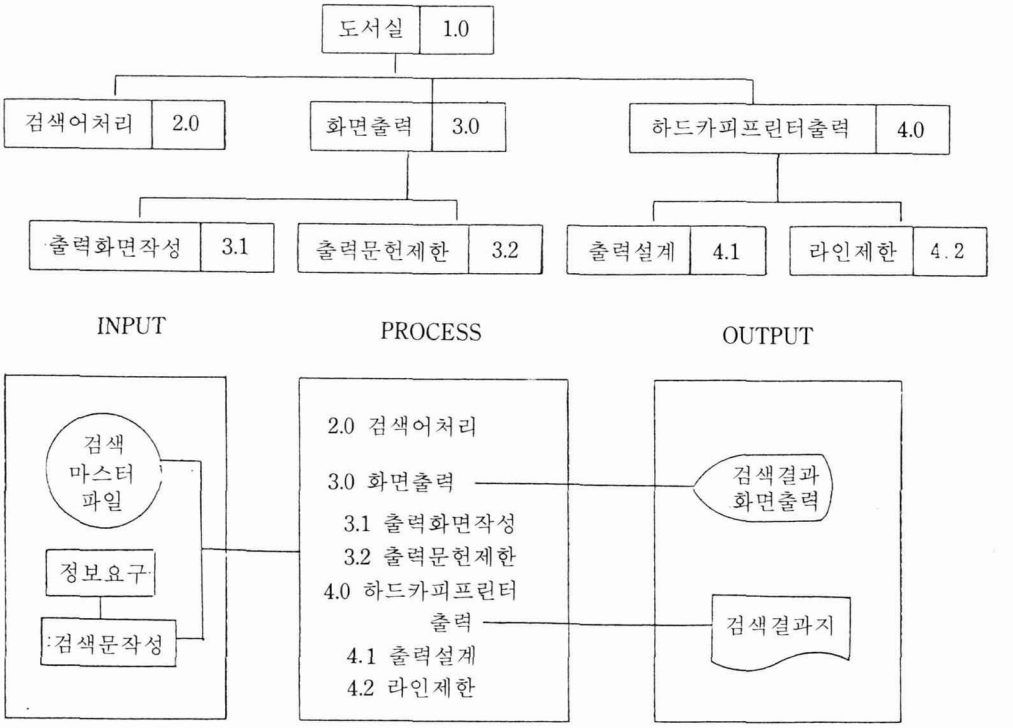

Fig. 5. HIPO of Output Design
출력설계의 HIPO

3. 파일설계(file design)

본 시스템에서 파일설계의 목적은 입력설계에서 만들어진 입력마스터파일을 분류하여 각 항목별로 트랜잭션파일을 만들고 오류를 수정한 후 검색에 필요한 도치파일, 색일파일 및 검색마스터파일을 만드는 것이다.

1) 마스터파일(master file)

(1) 입력마스터파일(input master file)

검색시스템에서 최초로 만들어지는 파일이며 앞서 언급한 입력설계에서 만들어진 12개의 항목으로 구성되어 있고 레코드의 구성과 항목의 길이는 〈그림 6〉과 같다.

CITATION NUMBER	AUTHOR 1	AUTHOR 2	AUTHOR 3	SUBJECT CODE	PUB-TYPE	CALL NUMBER	TITLE I	TITLE II	PLACE	RECORD N
I5	A20	A20	A20	A8	A1	A14	A150	A150	A4	

PUBLISHER	YEAR	PAGE	LANGUAGE	DESCRIPTOR 1	DESCRIPTOR 2		DESCRIPTOR 10
A35	A4	A7	A2	A30	A30		A30

Fig. 6. Organization of Input Master Record
입력마스터레코드의 구조

(2) 검색마스터파일(retrieval master file)

이 파일은 기억공간을 절약하고 화면의 1행(row)에 70칼럼(column)씩을 출력하기 위해 입력마스터파일을 재구성한 것으로 〈그림 7〉의 레코드구조와 같이 입력마스터파일의 일련번호 부터 청구번호까지를 1개의 레코드(70칼럼기준)로 하고 서명 I 부터 마지막 항목인 색인어까지의 길이를 계산하여 8개의 레코드(CONTENT 1~ CONTENT 8)로 만들었다.

2) 트랜잭션파일(transaction file)

입력마스터파일의 레코드를 각 항목별로 수정하기 위해 일시적으로 만들어지는 작업용 파일이다. 저자 및 색인어 트랜잭션파일은 한 문헌에 저자와 색인어와 2개이상 발생할 경우 열(column)로 구성된 필드를 행(row)으로 재정렬하여 구성하였다. 각 항목별 트랙잭션파일의 구조는 〈그림 8〉과 같다.

SEQ-NO	AUTHOR	BLANK	PUB-TYPE CALL-NO	CONTENT 1		CONTENT 8
A5	A45	3X	A17	A70		A70

Fig. 7. Organization of Retrieval Master Record
검색마스터레코드의 구조

RECORD	ORGANIZATION OF RECORD		RECORD	ORGANIZATION OF RECORD	
AUTHOR T-R	SEQ-NO	AUTHOR	PUBLISHER T-R	SEQ-NO	PUBLISHER
	A5	A20		A5	A35
SUB-CODE T-R	SEQ-NO	SUB-CODE	YEAR T-R	SEQ-NO	YEAR
	A5	A8		A5	A4
PUB-TYPE T-R	SEQ-NO	PUB-TYPE	PAGE T-R	SEQ-NO	PAGE
	A5	A1		A5	A7
CALL-NO T-R	SEQ-NO	CALL-NO	LANGUAGE T-R	SEQ-NO	LANGUAGE
	A5	A14		A5	A2
PLACE T-R	SEQ-NO	PLACE	DESCRIPTOR T-R	SEQ-NO	DESCRIPTOR
	A5	A4		A5	A30

Fig. 8. Organization of Transaction Records
트랜잭션레코드의 구조

3) 도치파일(inverted file)

(1) 색인어도치파일(decriptor inverted file)

색인어 검색을 위해 만들어지는 파일로서 입력된 색인어를 그대로 사용하는 색인어파일 (descriptor file)과 2개 이상의 단어로 구성된 색인어를 분리하여 색인어의 좌측절단과 양측절단 검색시 사용하기 위한 색인어절단파일(descriptor truncation file) 그리고 색인어 검색시 100건 이상 발생하는 색인어를 신속히 검색하기 위해 만들어진 2개의 클러스터파일(Clustered file : 서로 유사한 레코드들로 구성된 파일)로 구성된다. 이 파일들은 입력마스터파일에서 색인어와 일련번

호를 추출하여 순차적으로 정렬(sort)된 파일로 만든후 해당 색인어가 위치한 물리적인 레코드 위치인 블록주소(block address)와 그 블록 내에서의 위치를 나타내는 시작주소(start address) 그리고 색인어가 갖는 문헌건수를 계산하여 구성한 것이다.

(2) 저자도치파일(author inverted file)

저자 검색을 위한 파일로 입력마스터파일에서 일련번호와 저자필드를 추출하여 색인어도치파일과 동일한 방법으로 파일을 구성하였다.

(3) 서명도치파일(title inverted file)

본문의 제목을 검색하기 위해 입력마스터파일에서 일련번호와 서명필드를 추출하여 색인어도치파일과 동일한 방법으로 파일을 구성하였다. 또한 서명검색을 신속히 수행하기 위해 서명클러스터 파일을 만들어 사용하였다.

(4) 출판사도치파일(publisher inverted file)

출판사검색을 위해 입력마스터파일에서 일련번호와 출판사 필드를 추출하여 색인어도치파일과 동일한 방법으로 파일을 구성하였다.

4) 색인파일(index file)

(1) 색인어색인파일(descriptor index file)

앞서 언급한 색인어도치파일에 대한 모든 문헌들의 주소(address)를 기록한 파일로서 각 레코드는 블록단위(1 block : 180 column)로 구성되어 있으며 한 블록당 20개의 문헌주소를 기록하고 있다. 특히 이 파일은 색인어 검색시 해당문헌을 찾아 갈수있는 색인 (index)의 역할을 하는 것으로 색인어도치파일과 같이 4개의 파일로 구성되며 〈그림 10〉과 같이 모두 동일한 구조를 갖는다.

(2) 저자, 서명, 출판사색인파일(author, title, publisher index file)

저자, 서명, 출판사도치파일에 대한 문헌들의 주소를 기록한 파일로서 그 구성과 구조는 색인어색인파일과 동일하다.

RECORD	ORGANIZATION OF RECORD			
DESCRIPTOR INVERTED	DESCRIPTOR	BLOCK ADDRESS	STRAT ADDRESS	NO. OF BIBLIOGRAPHY
	A30	I5	I5	I5
AUTHOR INVERTED	AUTHOR	BLOCK ADDRESS	STRAT ADDRESS	NO. OF BIBLIOGRAPHY
	A30	I5	I5	I5
TITLE INVERTED	TITLE	BLOCK ADDRESS	STRAT ADDRESS	NO. OF BIBLIOGRAPHY
	A130	I5	I5	I5
PUBLISHER INVERTED	PUBLISHER	BLOCK ADDRESS	STRAT ADDRESS	NO. OF BIBLIOGRAPHY
	A40	I5	I5	I5

Fig. 9. Organization of Inverted Records
도치레코드의 구조

BIBLIOGRAPHIC ADDRESS 1	BIBLIOGRAPHIC ADDRESS 2		BIBLIOGRAPHIC ADDRESS 20
A9	A9		A9

Fig. 10. Organization of Descriptor Index Record
색인어색인레코드의 구조

5) 사전파일(dictionary file)

색인어검색시 참조하기 위한 파일로서 Medical Subject Headings와 Popline Thesaurus 그리고 사회복지용어집을 합쳐 국문과 영문을 따로 순차적으로 정렬한 국·영문 사전이며 입력마스터 파일을 작성할때 색인어 오류검사파일로도 사용되어진다. 레코드의 구조는 〈그림 11〉과 같다.

6) HIPO

입력, 처리과정, 출력을 그림으로 표시하면 〈그림 12〉와 같다.

KOREAN DESCRIPTOR	ENGLISH DESCRIPTOR
A40	A30

ENGLISH DESCRIPTOR	KOREAN DESCRIPTOR
A30	A40

Fig. 11. Organization of Dictionary Record
색인어사전레코드의 구조

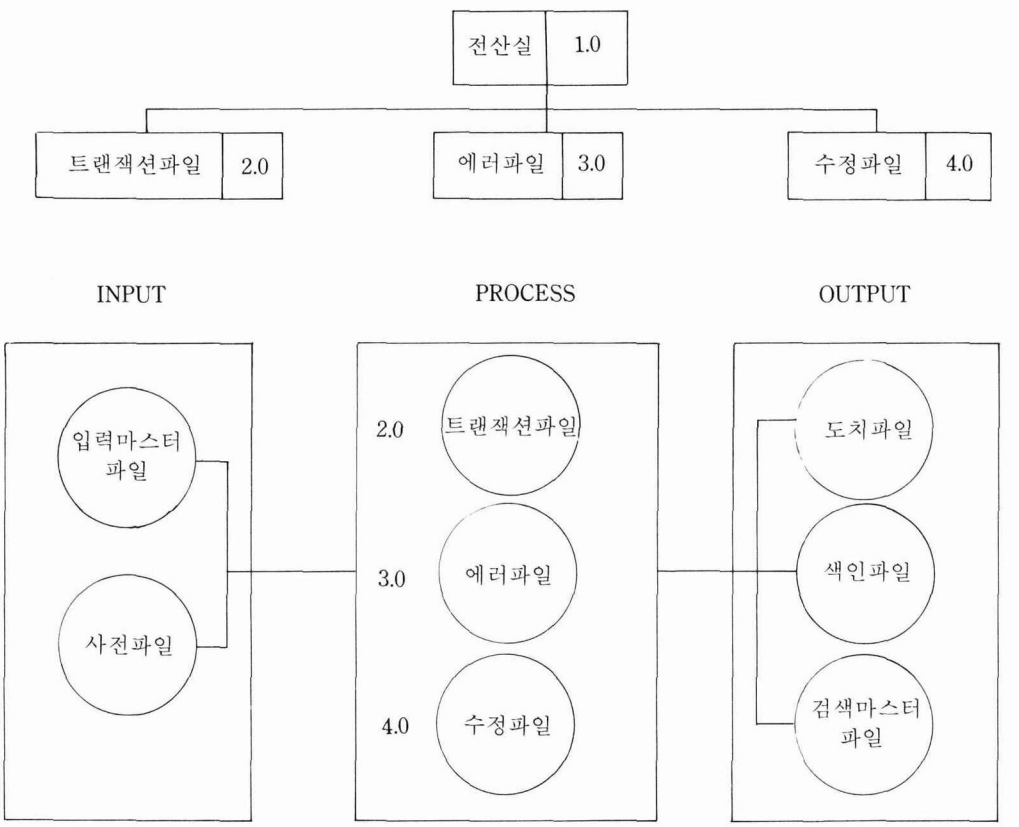

Fig. 12. HIPO of File Design
파일 설계의 HIPO

4. 프로세스설계(process design)

검색시스템에서 정보자료를 검색해 내기 위한 기준으로 채택한 기법을 검색기법이라 하며 검색기법으로는 이항검색, 불리안논리검색, 클러스터파일검색, 용어절단검색등이 있고 이러한 기법들은 검색시스템의 프로세스설계에서 사용되어진다.

1) 이항검색

순차적으로 정렬된 문헌을 처음부터 차례로 검색하지 않고 문헌의 중간부터 시작하여 검색어와 문헌이 갖는 키를 비교하여 그 결과가 크면 문헌의 상반부를 사용하고 작으면 문헌의 하반부를 사용하여 동일할때 까지 계속 시행하는 검색기법으로 이 방법은 매번 검색 때마다 문헌을 절반씩 나눔으로써 찾으려는 항목이 들어있는 범위를 단계적으로 좁혀 간다. 따라서 실제 찾고자하는 문헌이 매 검색후 반으로 줄어들기 때문에 총 문헌건수를 N으로 볼때 최대 $\log_2(N)$번의 검색이 필요하게 된다. 본 시스템에서 적용한 이항검색의 플로우차트는 〈그림 13〉과 같다.

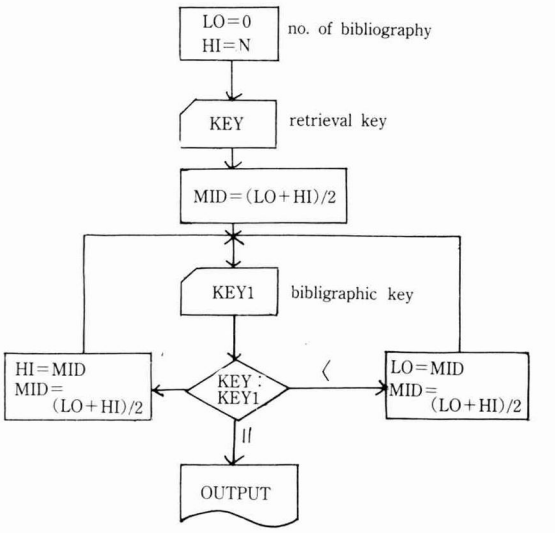

Fig. 13. Flow-chart for Binary Search
이항검색의 플로우차트

2) 불리안논리검색

불리안논리에 의한 검색은 현재 온라인 검색시스템에서 가장 많이 사용되는 검색기법으로 검색어와의 관계를 논리연산자인 AND, OR, NOT으로 표현하며 실제 온라인 시스템에서는 편의상 간단한 부호(*, +, .)를 많이 사용하고 있다. 본 시스템에서는 AND(*) 연산자를 사용하여 검색문을 작성하게 설계되어 있다.

```
HEALTH   317 1658  1784 19800 21500        결과
                                  RES    1784
PENSION 1564 1784 15678 18732
```

앞의 예와 같이 AND 연산자의 경우 317과 1564를 비교하여 PENSION쪽이 크므로 HEALTH의 다음 문헌번호를 비교한다. 1658을 만나면 HEALTH쪽이 크므로 1564의 다음것을 비교하게 되며 1784가 서로 만나면 최종기억장소인 RES에 1784를 옮긴다. PENSION의 끝번호인 18732가 자기보다 최초로 큰수인 19800을 만나면 19800 다음에 아무리 많은 문헌번호가 있더라도 더이상 처리하지 않고 끝내게 된다.

3) 용어의 절단검색

용어의 절단(truncation)이란 검색어로 사용하는 용어의 일부분을 생략하고 나머지 부분만을 검색어로 사용하는 기법으로 절단되는 위치에 따라 좌측절단(어두절단), 우측절단(어미절단), 양측절단, 중간절단으로 나눌 수 있다. 본 시스템에서는 좌측, 우측, 양측절단검색을 모두 사용할 수 있으며 우측절단 검색방법으로는 절단된 검색키의 길이 만큼 해당 문헌의 문자열을 읽어와 서로 비교하였고 좌측 및 양측절단검색은 두 단어 이상으로 구성된 색인어의 문자열 사이에 공란(space)이나 콤마(comma)가 있을 경우 한 단어씩 분리하여 새로운 파일로 구성하여 이 파일에서 검색어와 해당문헌을 서로 비교함으로써 처리가 가능하게 만들었다.

(1) 좌측절단검색

여러개의 다른 접두어를 갖는 관련용어 검색에 유용하다.

검색문	검색결과
%WELFARE	AGED WELFARE
	CHILD WELFARE
	FAMILY WELFARE
	SOCIAL WELFARE

(2) 우측절단검색

단수 복수 용어의 검색과 다양한 어미를 갖는 용어의 검색에 유용하다.

검색문	검색결과
FAMILY%	FAMILY FUNCTION
	FAMILY MEDICINE
	FAMILY ROLE
	FAMILY

(3) 양측절단검색

검색어의 앞쪽과 뒷쪽을 동시에 절단 함으로써 중간부분과 일치하는 용어를 검색할 수 있다.

검색문	검색결과
%PLANNING%	FAMILY PLANNING ASSISTANCE
	FAMILY PLANNING POLICY
	HEALTH PLANNING SUPPORT

4) 클러스터파일의 검색

유사한 내용의 문헌들을 같은 클러스터에 모아둠으로써 전체 문헌집단을 여러개의 클러스터로 나누어 검색하는 기법이다. 클러스터 별로 문헌레코드를 모아 조직한 파일을 클러스터파일이라 하며 각 클러스터는 해당 클러스터를 대표할 수 있는 키값을 갖게된다. 따라서 문헌검색시에는 전체문헌에서 찾지않고 검색키와 클러스터의 대표키를 비교하여 동일한 경우 해당 클러스터에서 문헌을 비교하여 처리한다. 본 시스템에서는 색인어와 서명의 검색에 사용하였다.

(1) 색인어 클러스터파일

본 시스템에서는 2개의 파일(색인어파일, 색인어절단파일)에서 100건이상 발생하는 유사 색인어를 추출하여 35개의 클러스터대표값을 주어 하나의 파일로 구성 하였다.

클러스터대표값	유사색인어
HOSTIPAL	HOSPITAL ADMINISTRATION
	HOSPITAL AUXILIARIES
	HOSPITALIZED HEALTH CENTER

(2) 서명 클러스터파일

국문문헌이 대부분이므로 서명을 가-하, 기타(영문포함)의 15개의 클러스터로 나누어 하나의 파일을 구성하였다.

V. 시스템평가

1. 시스템검색

모든 정보의 검색은 메뉴화면을 통하여 이루어지며 사용자는 CRT의 화면상에서"KIHASA"라는 명령어를 입력시키면 〈그림15〉과 같은 검색화면이 출력되면서 검색시스템과 연결이 된다.

만약 이용자가 이 시스템에 대하여 익숙하지 않을 경우 도움화면을 통하여 검색방법에 관한 정보를 얻을 수 있다. 검색방법을 익힌 사용자가 검색을 원할 경우 선택화면의 번호중 하나를 선택하여 다음과 같은 검색을 할 수 있다.

1) 검색화면 I 〈그림 16〉

색인어와 저자를 검색키로 사용하며 사용자는 최대 6개의 색인어에 대해 AND 검색을 할 수 있다. 본 시스템에서 AND 연산자는 "*" 를 사용 하였으며 명령어의 마지막은 "/"로 막아준다. 저자로만 검색을 원할 경우 색인어 부분에서 ENTER키를 입력하면 커서가 저자필드에 위치

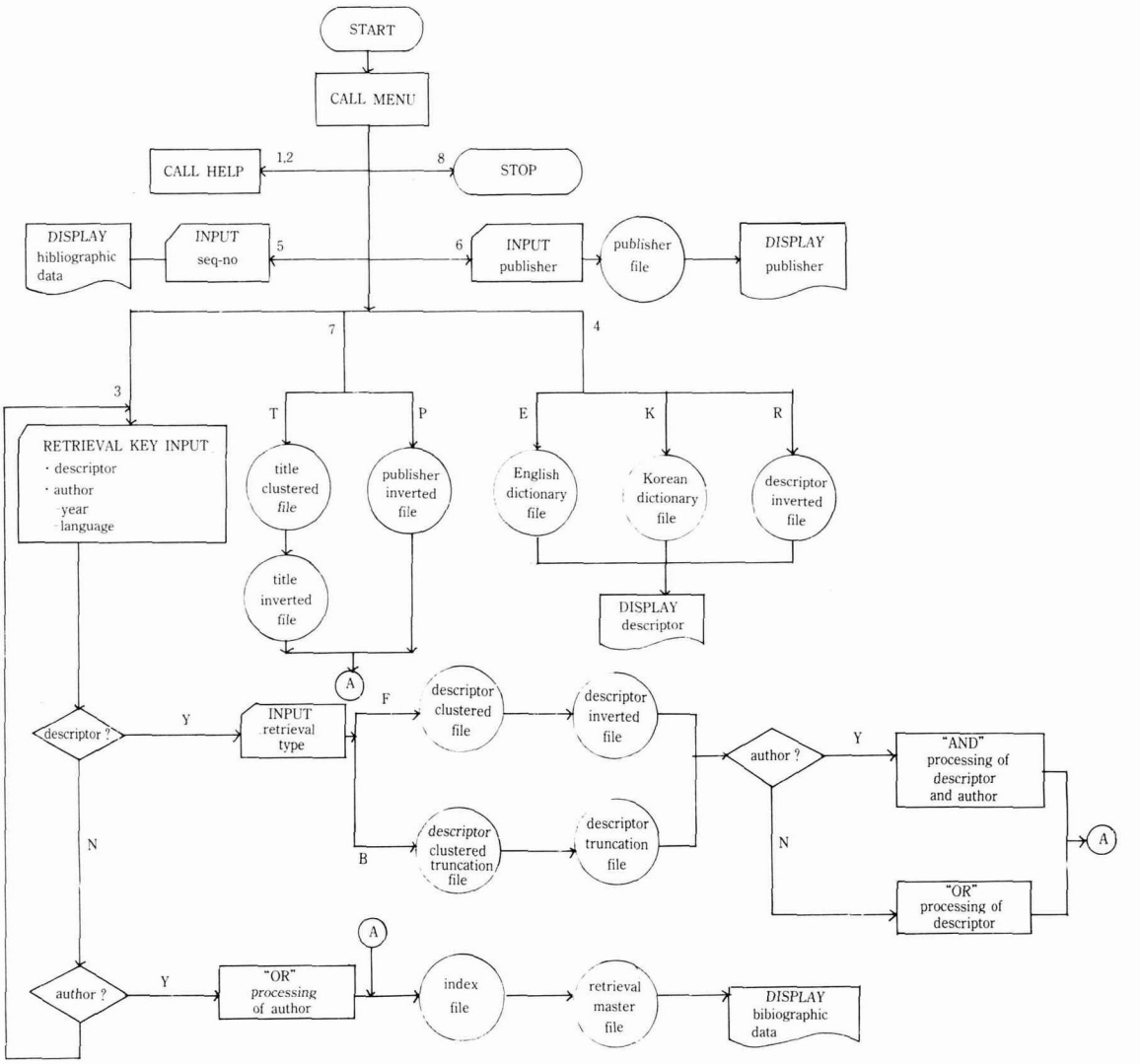

Fig. 14. Flow-chart of Retrieval Program
검색프로그램의 플로우차트

* * 선택 화면 * *

1. 한글도움화면　　2. 영문도움화면
3. 검색화면 I　　4. 색인어 출력(E,K,R) :
5. 일련번호 출력　　6. 출판사 출력
7. 검색화면 II(T : 서명 P : 출판사)　8. 끝

! 위 번호중 하나를 선택 하십시오 :

Fig. 15. Menu Screen for Retrieval
검색을 위한 선택하면

* * * 인구 · 보건 · 사회 문헌검색 화면 I * * *

A. 검색키

1) 색인어 :　　2) 저자 :

B. 제한사항

1) 년도 :　　2) 언어 :

검색을 원합니까 ?
(F : 우측검색, B : 좌측검색, R : 다시, ^ : 위)

Fig. 16. Screen I for Retrieval
검색화면 I

한다. 검색키는 항상 1개이상 존재해야 하며 모든 검색키는 연도와 언어를 제한사항으로 사용할 수 있다. 검색키 입력후 우측절단검색을 원할 경우에는 우측검색(F)키를 선택하고 좌측절단검색과 양측절단검색을 원할 경우에는 좌측검색(B)키를 선택할 수 있다. 사용자는 입력사항에 대한 오류가 발생했을 경우 R 혹은 ∧ 키를 이용하여 검색어를 수정할 수 있다.

2) 색인어출력

사용자가 해당 문헌에 대한 색인어를 모를 경우 E:영문, K:국문, R:원문키를 선택 함으로써 색인어 국. 영문 사전파일이나 색인어 파일에서 색인어를 추출하여 검색키로 사용할 수 있다.

3) 일련번호출력

원문의 일련번호를 알고 있을때 일련번호를 검색키로 사용하여 검색할 수 있다.

4) 출판사출력

출판사를 모를 경우 출판사파일의 출판사를 참조하여 출판사 검색을 할 수 있다.

5) 검색화면 II 〈그림 17〉

서명이나 출판사를 검색키로 사용할 경우 선택하는 화면으로 서명은 150칼럼, 출판사는 35칼럼을 사용하여 검색어를 작성할 수 있다. 색인어 검색과 같이 연도와 언어를 제한사항으로 사용한다.

```
*** 인구·보건·사회 문헌검색 화면II ***
A. 검색키
1) 서명 :          2) 출판사 :
B. 제한사항
1) 년도 :          2) 언어 :
! 검색을 원합니까? (Y/N, R : 다시, ^ : 위)
```

Fig. 17. Screen II for Retrieval
검색화면 II

2. 시스템평가

정보검색시스템의 평가는 일반적으로 검색효율, 신속성, 경제성의 세가지 측면에서 고려되어진다. 검색효율은 이용자와 직접적으로 관련된 평가기준으로 이용자의 만족도 및 정보서비스의 질을 측정하는 것이며 정보제공부서(도서실)에서 이용자와의 면담이나 설문조사를 통하여 측정이

Table 2. Distribution of Retrieval Key
검색키의 분포

(unit : %)

No. of Bibliography	Descriptor	Author	Publisher
1 ~ 5	62.1(2,608)	92.1(7,453)	73.9(955)
6 ~10	12.4(520)	4.6(369)	9.4(121)
11~20	9.6(405)	2.2(175)	6.6(85)
21~30	4.6(191)	0.5(39)	3.0(39)
31~40	2.5(105)	0.2(20)	1.6(21)
41~50	1.8(75)	0.1(11)	1.3(17)
51~100	3.6(152)	0.3(21)	2.3(30)
101~	3.4(141)	-(1)	1.9(25)
Mean	16.462	2.49	11.35
Maximum	1,032	133	664
Sum	69,091	20,219	14,682

Table 3. Turn Around Time of Retrieval Key
검색키의 응답시간

Retrieval Key		No. of Bibliogrphy	Turn Around Time (second)
Descriptor	Maximum	1,032	2
	Mean	17	1
Author	Maximum	133	1
	Mean	3	1
Publisher	Maximum	664	3
	Mean	12	1
Title	Maximum	8	1
	Mean	1	1

가능하다.

신속성과 경제성은 정보검색작업에 소요되는 시간과 경비를 측정하는 것으로 주로 시스템운영자의 입장에서 관심을 갖는 평가기준이다.

본고에서는 시스템개발부서에서 측정이 가능한 신속성 즉 검색소요시간을 시스템의 평가기준으로 사용하였다

검색소요시간(응답시간)을 측정하기 위해 검색용 파일에서 〈표 2〉와 같이 한 검색키가 보유하는 문헌건수에 관한 분포를 구하고 각 파일에서 최대 및 평균값을 갖는 키를 찾아서 그 키에 대한 검색을 수행한 결과 〈표 3〉과 같은 응답결과가 나왔다.

〈표 3〉에서와 같이 대부분의 검색에서 응답시간은 대기시간이 거의없는 1-2초로 나타났으며 출판사 검색시 다소 지연된 것은(응답시간 3초) 검색속도를 빠르게 하기위한 출판사 클러스터파일을 구성하지않았기 때문이다.

VI. 결 론

본 인구. 보건. 사회 문헌정보 검색시스템은 한국보건사회연구원도서실이 수집, 관리하고 있는 관련분야 문헌을 연구자가 컴퓨터를 이용하여 색인어, 저자, 출판사, 서명으로 검색하기 위하여 개발한 것으로 사용자가 간단하고 쉽게 검색할 수 있게하고 검색시 소요되는 응답시간을 줄이는데 개발의 중점을 두었다. 따라서 이용자의 편의를 위한 Man-Machine Interface를 좋게 하기 위해서 모든 검색은 메뉴화면을 통하여 할 수 있게 하였고 검색소요시간을 최소화하기 위하여 클러스파일을 사용하였다. 본 시스템의 특징을 보면 다음과 같다. 첫째, 색인어는 절단(우측, 좌측, 양측)검색과 불리안논연산자(Boolean Logic operator)인AND(*)를 이용하여 검색할 수 있다. 둘째, 출판사와 서명은 우측절단하여 검색이 가능하다. 셋째, 문헌의 일련번호로 검색할 수 있다. 넷째, 색인어 및 출판사를 모를 경우 색인어 사전파일이나 원문에서 해당 색인어를 참조하여 검색키로 사용할 수 있다. 다섯째, 모든 검색시 연도와 언어는 제한사항으로 사용된다.

본 시스템에 사용한 통제어휘집은 보건, 사회분야의 완벽한 어휘집이 아니기 때문에 해당분야의 전문가를 통하여 많은 수정 및 보완이 이루어 져야하며 연구자에게 보다 상세한 내용의

정보를 주기 위해 입력항목에서 빠진 초록의 추가 입력에 관한 문제가 논의 되어 져야한다.

현재 본 시스템의 갱신(update)주기는 1년이기 때문에 연구자에게 최신의 정보를 제공해 줄 수 없다. 따라서 update기간을 최대한 줄이기 위해서는 전문인력 및 적절한 예산이 뒷받침 되어 져야할 것이다.

또한 정보검색시스템은 정보의 내용분석과 가공을 담당하는 정보전문가와 시스템설계를 담당하는 컴퓨터전문가의 협력으로 이루어지기 때문에 시스템의 질을 높이기 위해서는 상호간의 협력과 노력이 요구된다. 그리고 본 시스템은 정보관리 시스템의 일부분인 정보검색시스템만 개발되어 졌기 때문에 필요한 데이타를 추가로 입력하여 수서, 대출, 목록시스템과 같은 서브시스템과 상호 통합하여 사용할 수 있도록 토탈시스템구성이 이루어져야 할 것이다.

참 고 문 헌

산업정보기술원, 정보관리교재 시스템사례편, 서울, 1991.

송태민, "인구 및 보건 문헌정보 검색시스템 개발에 관한 연구", 인구보건논집, 제6권 제1호, 1986, PP. 118-131.

양해술, 시스템분석과 설계, 상조사, 서울 1985.

이기식, 정왕호, 소프트웨어 생산기술, 정익사, 서울, 1989.

이석호, 화일처리론, 정익사, 서울, 1985.

이석호, 데이타베이스론, 정익사, 서울, 1987.

정영미, 정보검색론, 정음사, 서울, 1987.

한국에너지연구소, 원자력관계 기술보고서 검색시스템, 서울, 1983.

C. J. DATE, An Introduction to Database System, London, 1981.

John G. DONOVAN, System Programming, MCG-RAW-HILL Book Company, New York, 1972.

〈**Summary**〉

Development of Bibliographic Information Retrieval System on Population, Health and Social Affairs

Tae-Min Song*

This study develops a new bibliographic information retrieval system, using FORTRAN-77 as the program language and MV 15000-8 Super Mini Computer as hardware. In 1986 the Korea Institute for Health and Social Affairs(KIHASA) constructed a bibliographic information retrieval system. It contains descriptors, author, title, call number, subject code, publication type, place, publisher, year, page, language, and citation number from books, journal articles, conference papers, and statistical materials. The system contains approximately 15,000 pieces of bibliographic data on population, health, and social affairs and about 2,500 pieces of bibliographic data are being added every year.

How can the information retrieval system be maintained and repaired easily? This has been one of the most crucial issues for the managers. In developing a new bibliographic information retrieval system, the flow-chart and HIPO(Hierarchy Plus Input-Process-Output) are employed as diagraming tools for easy maintenance and repair of the system. The menu screen is also used to improve man-machine interface, and the clustered file to minimize turn around time.

The retrieval techniques constructed in the system are : (1) binary search, (2) boolean logic retrieval, (3) truncation retrieval, and (4) clustered file retrieval. The binary search is an iterative procedure that can be used in retrieving the bibliographic data indicated by the retrieval key. The iterative procedure is a technique of successive comparisons between retrieval and bibliographic keys. The initial comparison begins at the midpoint of the bibliographic data. The successive comparisons are made in the upper or lower half sections of the bibliographic keys, and approaches the same value as the retrieval key.

Boolean logic retrieval is a technique in which the relationship between bibliographic and retrieval keys is expressed in such logic operators as "and', 'or', and 'not'. The operator 'and' is employed in this system. The truncation retrieval is a technique which does not use to whole retireval key, but only a part of it. The retrieval key can be truncated on the right side, the left side, or both sides in the new system. The clustered file retrieval is a technique which clusters the bibliographic data according to their contents, and compares the bibliographic and retrieval keys in the clustered files. The new system employs the clusterd file technique in retrieving descriptor

* Senior Computer Programmer, Korea Institute for Health and Social Affairs.

and title.

The keys used in the new bibliographic information retrieval system are : descriptors or truncated descriptors at right, left or both sides ; boolean logic operator 'and' in descriptor ; title or right-truncated publisher ; author or combined key of descriptor and author. Year and language are used as restricting keys in the retrieval procedures.