

BRS/SEARCH 정보검색 시스템이란?

What is BRS/ Search System?



서진이
산업기술정보원 DB사업부
Seo, jin-ry,
Korea Institute Of Industry
& Technology Information

현대는 '정보의 홍수'라 불리울 만큼 우리는 폭증하는 정보속에서 생활하고 있다. 의미있는 데이터의 집합인 정보, 데이터베이스의 양이 많아지고, 복잡, 다양해짐에 따라 정보의 관리가 점차 어려워지게 되었다. 그러한 수요와 요구에 따라 데이터베이스 관리체제, 자료관리 체제가 등장하게 되었다.

IRS 와 DBMS

정보검색시스템(IRS)과 데이터베이스관리 시스템(DBMS)은 이용자의 검색, 갱신요구가 실시간(Real time)으로 처리되며, 시스템의 장애 등 고장시 데이터의 복구기능이 있으며, 사용·이용자별, 데이터베이스별, 항목별 등의 보안유지 기능을 가지는 편리한 관리체제이다.

DBMS는 기존의 단순한 IRS 보다는 관리에 중점을 두었고, 데이터의 중복방지를 하여 중복된 데이터로 인해 메모리 공간을 차지하게 하지 않

아 메모리 공간도 절약해주는 이점이 있다.

반면 IRS는 검색위주의 시스템이므로 검색방법이 DBMS보다 복잡, 다양하며, 검색을 위한 키워드 추출 기능이 강화되어 있다. DBMS 전문회사의 성장세가 빠른것을 보면 알 수 있듯이, 시중에는 Oracle, Informix, Sybase, Ingres, UniSQL 등의 여러 DBMS가 상품화 되어 있다. 국내 최대의 DBMS판매량 및 사용자층을 보유하고 있는 Oracle DBMS는 타 DBMS 보다 메인 메모리를 많이 소요한다는 단점이 있으나 반면에 개발 툴을 많이 가지고 있다. Sybase는 클라이언트-서버 기능위주로 많은 호평을 받고 있으며, 이미지 처리 위주인 Uni-SQL과 분산처리에 강한 Ingres등이 주류를 이룬다.

그러나 시중의 외국산 DBMS는 보다 더 완벽한 한글화와 국산화의 지속적인 개발이 필요하다. 검색 위주의 IRS는 STAIRS, HAIRS, Search Manager/370, BRS/SEARCH, FAIRS-I,

BASIS-plus등이 상품화 되어 있어, 수백만건의 많은 정보를, 많은 이용자가 동시에 검색할 수 있게 한다. 온라인 정보 서비스를 위해선 검색위주의 IRS가 더 바람직하다.

BRS/SEARCH

1982년에 국내 최초로 온라인 정보서비스를 실시한 산업기술정보원(KINITI)에서는 IBM 호스트하에서 가동되는 BRS/SEARCH 라는 IRS시스템을 채택하여 온라인 정보서비스를 실시하고 있다.

BRS와 STAIRS

BRS/SEARCH (Bibliographic Retrieval Service)는 세계적인 미국의 BRS/ONLINE社에서 개발, 운용하고 있는 IRS이다. KINITI에서는 기존에는 IBM 호스트하의영문정보 전용의 'STAIRS'와 STAIRS를 기초로 한글화 시킨 'KIROS'를 운용하였다.

BRS/SEARCH는 STAIRS를 기본으로 개발되었기에 명령어가 STAIRS와 유사하며 데이터베이스의 구조도 유사하다. 또한 1,2 바이트의 혼용 문자검색이 가능하여 영문, 한글을 같이 검색할 수 있는 통합관리 시스템이다. 검색속도나 검색기능(논리 연산 AND, OR, ADJ, NOT기능, 좌우절단, 항목제한 검색기능, 멀티 데이터베이스, 다양한 화면 출력형태 지정)이 탁월하다. 또한 커

<표1> DBMS와 IRS의 비교

	DBMS	IRS
data구조	복잡다단	단순구조
data항목수	다수개	소수개
data종류	비정형 데이터	정형 데이터
갱신방법	수시 갱신	일괄 갱신
data중복성	중복방지 가능	중복방지의 어려움
적합한 분야	통계, 인사정보등 데이터 항목과 이용자요구 형태가 많은 경우	문헌정보처럼데이터 항목이 비교적 적고, 이용자의 요구형태가 일정한 경우

다란 특이점은 데이터베이스를 관리, 구축하는데 있어, 데이터 압축(Compression)기능이 있어 디스크 공간의 절약에 일조를 한다는 점이다. BRS/SEARCH는 통계 데이터와 같은 비정형 데이터가 아닌 정형화된 Full-text 검색에 기초를 둔 전자문서 관리시스템이다. 나아가 2진 데이터, 오디오, 비디오, 이미지, 그래픽 데이터 등 하이퍼 멀티미디어의 정보를 액세스 한다.

BRS 데이터베이스

BRS 데이터베이스는 하나 하나의 문자가 모여서 된 단어(Word)와 각각의 스페이스로 구분되어지는 단어와 단어끼리 모여서 된 문장(Sentence)과 문장의 집합인 항목(Paragraph)과 여러 항목들이 모여서 된 문서들로 나뉘어진다. BRS는 IRS이므로 DBMS와는 달리 관계형데이터베이스나 계층형 데이터베이스로 데이터가 관리되지 않고 Full Text 화일, 역변화 파일로 데이터가 관리되어 진다. BRS의 모든데이터베이스는

<표2> STAIRS/ BRS 명령어 비교

기 능	BRS	STAIRS
검색	..Search	..Search
제한 검색	..Limit	..Select
화면 출력	..Print	..Browse
오프라인 출력	..PrintOff	..Mail
검색 종료	..Off	..Off
DB교체	..Change	..Change
검색식 저장	..SaVe	..Save
검색식 실행	..Exec	..Exec
소트	..Sort	..Sort
검색식 삭제	..Purge	..Purge
검색식 화면 출력	..Display	..Display
도움말	..What	..Help
오프라인 출력 병합	..Merge	-
데이터 우선 제한	..Universe	-
멀티 DB	..Crosin ..Crosout	-
Spell 체크	..Set alpha	-
영미의 철자 다른	..Set	-
단어 체크	medspell	-
단.복수 체크	..Set plurals	-

정보검색을 가능케 하는 검색파일과 보조파일로 관리가된다. 검색을 위한 파일은 본문파일(Text File), 본문색인 파일(Text Index File), 변환파일(inverted file), 사전파일(Dictionary File), 역사전파일(reversed dictionary file)들이다. 본문 파일은 검색된 결과를 화면에 비추거나 용지에 출력할때 불러지는 파일로 실제 입력되거나 도입된 데이터 소스를 BRS에서 인식할 수 있는 형태로 바꾼 형태의 파일과 같다.

본문색인 파일은 본문파일내의 각 문헌의 각 시작점을 가리키는 포인터와 검색된 결과에 제한을 하는 항목의 모든 정보가 수록되며, 각 문헌마다 한 레코드씩 만들어진다. 변환 파일은 검색을 위한 중요한 색인어 파일로 데이터베이스에 단어가 나타나는 문헌의 수와 단어의 총 출현 회수 등이 수록되며 각 단어에 대한 정보는 출현 단어의 대소문자 표시, 상대적인 문헌번호, 항목코드, 문장번호, 단어번호 등을 갖고 있어 검색 질문식에 입력된 논리 연산자에 따른 결과를 얻기 위해 사용된다. 사전파일은 데이터베이스내에 나타나는 모든 단어와 단어의 출현회수, 문헌수를 수록하고 있으며, 단말기를 통해 입력되는 중요한 키워드의 출현 회수와 문헌수를 이 파일로 읽어 곧바로이용자에게 알려준다.

역사전파일(Reversed Dictionary File)은 사전 파일의 각 단어들의 철자배치를 거꾸로 하여 순차배열된 파일로 좌절단이나 양측 절단 검색시 필요하다. BRS/SEARCH 내에서의 문헌정보가 가능하기 위해서 지금까지 기술한 검색파일 뿐만 아니라 여러종의 보조파일이 필요하다. 데이터베이스 제어파일 (DataBase Control file), 제한필드 파일(Limit Field File), 이용자 관리를 위한 이용자 등록파일(이용자 마스타 파일, 이용자 부록 파일), 실제 이용자들이 검색한 질문식 저장파일(Query file), 이용자 사용 시간파일(Account file), 메시지파일(Message file), 오프라인 프린

<표3> BRS의 데이터베이스 구성

000054020000	문서1 (전체55 건)
DOCID	0005401-BIST94E22671 0001
자료형태	T
기사구분	A1
사용언어	KO
제작년월	9501
제목해의 과학 기술 정보 데이터베이스 구축에 관한 연구	
저지한국과학기술연구원연구개발정보센터	
서지사항	연구보고서 (KOR) 914.9678.94-624 P123p 1994;
KIRP-	
분류	ET0610
초록	과학 기술 정보는 중요한 국가 지원으로 인식되고 있으며 연구 개발 활동 및 과학기술 정책결정 과정에서 필수적인 요소이다. 해외의 첨단기술 정보의 조직적 수집과 활용능력을 강화하고 기술 보호주의에 능동적으로 대처해야 한다. 이를 위해 수집된 정보를 용이하게 활용할 수 있도록 데이터베이스를 구축해야하며, 또한 정보 이용자의 요구를 분석하여 정보의 수집에 활용하는 체제를 갖추어야 한다. 본 연구를 수행함에 있어다음과 같은 목표를 설정하고 연구개발 하였다.
색인어	DB, 데이터베이스, 정보가공, 정보검색, 정보처리

트 명령시 관련되는 오프라인 프린터 제어파일 등이 관련 보조파일 들이다.

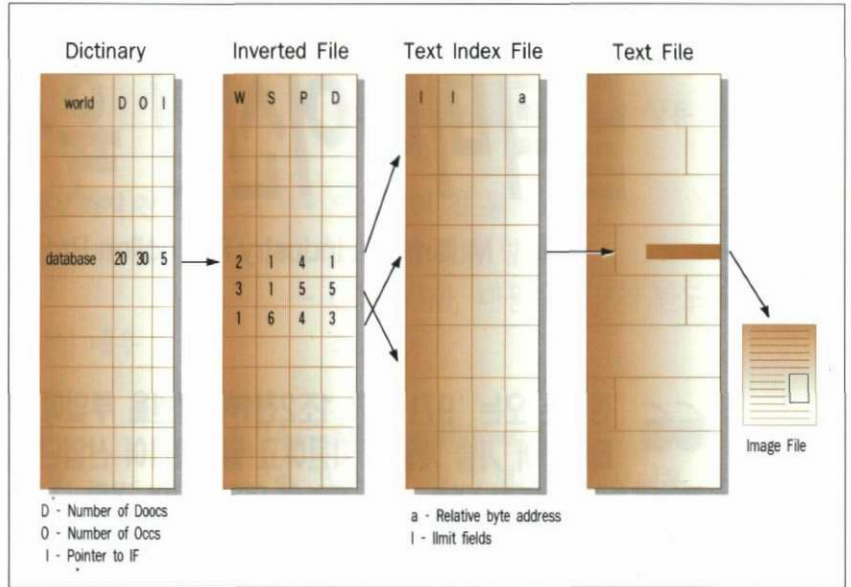
BRS 데이터베이스 구축

정보들이 이용자가 원하는 대로 검색이 되게 하기 위해서는 데이터베이스를 구축하는 일련의 과정이 요구된다. 형식 변환과정이 우선된다. 특정한 표준형식에 따라만들어져 교환되는 컴퓨터 가독형 자료와 직접수집, 가공, 제작되는 자료를 BRS에서 인식할 수 있는 자료의 형태로 변환하는 포맷 및 코드의 변환 과정이다.

포맷변환된 파일을 가지고 검색에 가장 중요한 인덱싱 프로세스를 거쳐야 한다. 이 과정에서 단어 그 자체로는 아무런 의미를 가지지 않는 대명사, 전치사, 접속사, 관사 등을 집합으로 만든 불용어 파일을 참조하여, 검색 색인어 추출시 본문파일내에는 존재하여 화면에는 보이더라도 색인파일에는 들어가지 않도록 한다. 또한 한글 데이터

처리시 조사 테이블이 있어 조사 처리를 가능하게 하고, 한글/한자 변환 테이블이 있어서 한글, 한자의 처리를 한다. 구축하고자 하는 데이터베이스의 내용 및 특성을 잘 이해하며 그 데이터베이스마다의 구축 과정을 정립할 필요가 있다. 소스 데이터를 읽어 본문파일과 본문색인파일을 만들며, 입력 데이터에 있는 모든 단어를 추출하여 불용어는 버리고, 중요한 구(句), 단어마다 문헌번호, 항목코드, 문장번호, 단어번호 등을 붙여서 단어파일을 만든다.

〈표4〉 BRS의 데이터베이스 구조



모든 단어를 순차배열로 정렬하여 분류된 단어 파일이 생긴다. 분류된 단어를 같은 단어이면 한 단어만 남기고 같은 단어의 포인터 부분은 계속 묶어 나간다. 분석된 단어파일을 읽어 사전파일을 만들고 포인터 부분을 읽어 변환파일이 생성된다.

일련의 과정을 거쳐서 데이터베이스가 구축되어 비로소 이용자가 정보를 검색할 수 있다. 정보 검색시스템인 BRS는 IBM의 온라인 시스템 CICS하에서 구동되어진다. DEMON을 통한 데이터의 입력, ODDM을 통한 데이터베이스 정의, 수정, 등록 및 데이터베이스 구축을 통해 명령어 (Command) 방식과 메뉴(Menu)방식의 두가지 방식으로 정보를 검색하게 한다. 더 나은 시스템을 위해선 BRS/Thesaurus 및 여러 유저 인터페이스의 개발이 필요하다.

지금까지 BRS/SEARCH에 관해 간단히 기술하였는데, 효율성의 극대화를 위해선 보다 더 적절한관리 및 데이터베이스의 보완, 다른 시스템과의 연계를 위한 Tool의 개발 등 유지관리가 계속적으로 필요하다. DC

〈표5〉 BRS 데이터베이스 구축 과정

