

# 시장 선점 향한 사활전 수주전 전개

인터넷의 발전함에 따라 정보검색이 대중화되면서 정보검색시스템 시장이 춘추전국시대를 맞이하고 있다. 국내 정보검색시스템과 외국 정보검색시스템 업체가 공공기관 및 전자도서관을 중심으로 치열한 경쟁을 벌이고 있다. 정보검색시스템의 시장동향 및 제품별 특징에 대해 살펴봤다. <박민식 기자>

## 정보사회의 필수요소

현대사회를 한마디로 표현한다면 정보사회라 할 수 있다. 정보사회의 가장 큰 특징은 정보의 디지털화와 급속한 양적 팽창이라고 할 수 있다. 이러한 흐름은 흔히 정보의 홍수라는 말로 대변되어 왔다. 하지만 모든 정보가 다 유용한 것은 아니며 필요한 정보를 효과적으로 추출, 활용할 때 그 정보는 빛을 발하는 것이다. 이러한 시대적 요구에 의해 탄생된 것이 정보검색시스템이다.

정보검색시스템은 수집된 각종 정보 데이터들의 내용을 자동으로 분석하여 자체의 데이터베이스를 구축하고 사용자 요구에 따라 필요한 정보를 고속으로 검색해 주는 시스템을 말한다. 다시 말해 수집된 정보를 찾기 쉽고 관리가 용이한 형태로 가공하여 사용자 요구가 있을 때 해당

정보를 쉽고 빠르게 제공하는 시스템이다.

초기 정보검색시스템은 도서관에서 책을 찾는 방식과 마찬가지로 누군가가 수작업으로 책의 저자와 출판사, 제목 등을 입력하고 이용자로 하여금 찾아볼 수 있도록 하였으며 이러한 방식은 이용자가 제목이나 저자 등을 정확히 알고 있어야만 찾을 수 있었다. 예를 들어 '도서관'에 대한 자료를 얻고자 할 경우 이용자가 할 수 있는 방법은 '도서관'이라는 단어가 제목에 있는가를 찾거나 백과사전을 빌리는 정도였다.

그러나 현재의 정보검색시스템은 문헌의 본문에서 색인어를 추출하므로 '도서관'이라는 단어를 입력하는 것으로 '도서관'을 언급한 모든 문서를 찾아 주며, 조건을 부여하는 정보에 따라 문서 범위를 제한할 수

도 있다.

이러한 정보검색시스템은 기본적으로 두 요소에 의해서 성능이 좌우되는데 색인과 검색 모델이 그것이다. 검색모델의 경우 학교나 연구기관을 중심으로 실험적인 방식이 사용되고 있으며 상용에서도 거의 유사한 방식이 사용되고 있다.

하지만 색인의 경우는 언어마다의 특징이 있어 언어별로 색인방식을 달리해야만 한다. 한글의 경우 철자 검사기를 기반으로 자동 색인기가 발전하여 최근에는 자동색인기 분야에 대한 관심이 높아지고 있다. 또한 색인 작업의 효율을 높이기 위하여 통제어를 지정하거나 시소러스, 동의어 등을 구축하고자 하는 노력이 계속되고 있어 자동색인 분야에 대한 발전을 더욱 가속화될 전망이다.

정보검색 시스템은 얼핏 생소하게

느껴질 수도 있지만 사실상 우리의 실생활에 아주 밀접하게 여러 분야에서 응용되고 있다.

## 국내 정보검색시스템 시장 동향

국내 정보검색시스템 시장은 몇년 간의 도입단계를 걸쳐 공공기관, 연구소, 신문사, 방송국 등 대량의 데이터베이스가 구축되어 있는 기관들을 중심으로 형성되어 왔다. 특히 현재는 각 대학들의 가상대학 및 전자도서관 구축 열풍에 따라 전체 시스템 상에서 정보검색시스템은 필수요소로 자리매김했다. 이같은 시장 상황에 의해 대학은 정보검색시스템 공급업체에게는 시장 활성화를 위한 인프라가 되고 있다.

현재 국내에 공급중인 정보검색시스템은 크게 국내 기술로 개발된 제품과 외산 검색엔진을 한글화한 제품으로 분류할 수 있다.

다센의 다센 21, 쓰리소프트의 인포에이스, 한국정보공학의 레이다, 소프트와이즈의 SOFTBOT, 엔피아 시스템의 Spider, 데이콤의 심마니 등이 국내 기술로 개발된 제품들이다.

외산 검색 엔진을 응용하거나 한글화한 제품으로는 사이버코리아의 닥터서치, SK컴퓨터통신의 BRS Search, 쓰리소프트의 서치 97, 포도시스템의 Basis plus, 나눔기술의 컴파스, 창신컴퓨터의 ReQuest, 한전정보네트웍의 U-reka, 화상검색시스템인 Excaliber사의 Visual Retrievalware 등이 있다.

정보검색시스템 공급업체들의 영업 전략은 다양하지만 크게 세가지

로 나누어 볼 수 있다.

첫째가 여타의 인터넷 서비스와 같은 형태로 사용자들에게 인터넷상에서 서비스를 제공하는 방향이다. 둘째는 별도의 에이전트 시스템을 개발하여 특정분야에 대해 전문 버전을 제공하는 전략이다. 이는 맞춤 서비스라고 할 수 있다. 셋째는 애플리케이션 형태로 개발하여 사용자가 자유롭게 정보검색 서비스를 제공하는 것이다.

## 업체 및 제품 현황

### 쓰리소프트 - 인포에이스, 서치 97

쓰리소프트는 자체 개발한 인포에이스와 美 베리티(Verity)사의 검색엔진을 도입한 서치 97을 공급하고 있다.

인포에이스는 자체 개발한 정보검색시스템으로 클라이언트/서버는 물론 인터넷 및 인트라넷 상에서 영문뿐만 아니라 한글까지 검색이 가능하도록 처리된 전문(全文) 검색엔진이다. 자체 디자인한 화일 시스템을 이용하여 검색 속도가 빠르며 각종 필터를 사용하여 HTML문서 뿐만 아니라 각종 워드프로세서, 표계산 프로그램 및 프리젠테이션용 프로그램의 문서들까지도 검색할 수 있다.

한편 美 Verity사의 검색엔진을 도입한 서치 97은 인터넷/인트라넷 상의 정보 검색이 가능하도록 웹브라우저 기반의 사용자 인터페이스를 채택한 것이 특징이다. 구성은 디벨로퍼 키트, 인포메이드 서버, 에이전

트 서버, 인트라넷 스파이더 등 4개의 제품으로 구성되어 있다.

구축 사이트에는 정보통신정책연구원, 한국방송광고공사, 중앙일보, 문화일보, 경향신문, 국회도서관, EBS, 국방과학연구소 등이 있다.

### 다센 - 다센21

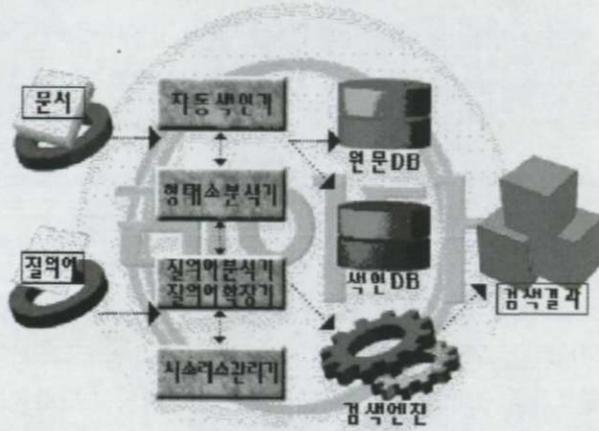
다센이 자체 개발한 다센21은 검색어에 대한 형태소 분석을 실시해 정확률 및 편의성을 부여했으며 다국어 동시검색, 시소러스 참조검색 등 지식베이스 기반의 검색을 제공한다. 풍부한 경험축적에 의한 안정화된 정보 검색시스템을 제공하며 형태분석기, 검색엔진, 웹 로봇 등으로 구성되어 있다.

순수 국내기술로 개발되어 지난 94년부터 공급하고 있으며 구축 사이트에는 법제처, 토지공사, 석유개발공사, 불교CATV, 부산방송, 포스코, 한국소비자보호원 등이 있다.

### 한국정보공학 - 레이다

한국정보공학이 자체 개발한 레이다 정보검색시스템은 대용량 정보를 효율적으로 관리할 수 있는 정보관리 기능과 시소러스 관리기능 등을 갖추고 있다. 구성은 자동색인기, 정보검색엔진, 질의해석기, 문서변환기, 정보관리기, 시소러스 관리기 등으로 구성되어 있다.

자동색인기는 문서정보전달, 단어정보추출, 어근추출, 불용어처리를 수행하며 문서변환기는 다양한 형식의 문서들을 레이다 엔진에서 인식



〈그림〉 레이다의 서비스 흐름도

할 수 있는 형태로 변환시켜 준다. 질의해석기는 사용자가 입력한 질의보다 정확하고 섬세한 검색결과를 얻기 위해 사용된다. 또한 질의로부터 검색결과를 얻어 내는 기능을 하는 것이 검색엔진이다. 정형데이터베이스와 비정형 데이터베이스로부터 자료검색을 수행하고 검색결과에 대한 부수적인 처리를 수행한다.

레이다는 SGML 등 다양한 문서 포맷을 지원하며 서로 다른 색인 데이터베이스간 통합 검색이 가능하다. 현재 인터넷용으로 검색시스템을 공급하고 있으며 구축사이트에는 전자신문사, 학술진흥재단, 석유개발공사, 한국전산원, 전자통신연구원 등이 있다.

**소프트와이즈 - SOFTBOT**

소프트와이즈가 자체 개발해 94년부터 공급하고 있는 SOFTBOT은 자연어 검색이 가능한 비정형 텍스트 DBMS로 정보의 관련성을 자동 계산해 주는 전문가 시스템을 내

장하고 있다.

자료의 전문(全文)을 색인화하여 문서 내의 모든 단어를 검색할 수 있는 전문 검색 엔진과 말로 질문하듯이 자연스러운 문구를 입력하면 엔진의 지능적 질의어 해석기에서 자동으로 해석하여 검색하는 자연어 검색 시스템을 갖추고 있다.

또한 SOFTBOT은 검색된 정보와 질의어와의 관련성을 통계적인 방법으로 자동 계산, 관련성이 높은 정보순으로 사용자에게 제시한다. 구축 사이트로는 대법관 도서관, 조흥경제연구소, 삼성경제연구소, 삼성자동차, 경찰청, 디지털 조선일보 등이 있다.

**현인정보 - 마이다스**

현인정보에서 개발한 마이다스는 순수 국내 기술로 개발된 자연어/전문 정보검색 시스템이다. 마이다스는 한글형태소 분석기와 자동색인 시스템에는 양방향 최장일치법을, 영문에서는 불용어 선처리법을 적용

함으로써 형태소 분석과 색인시 Hard Disk Access Time을 기존 제품보다 47%~82%까지 감소시켜 전체 시스템 속도를 향상시킨다. Sybase, Oracle, Informix, UniSQL 등 기존의 모든 DBMS와 자유롭게 연동할 수 있다.

또한 MS Office, HWP, 훈민정음, ASCII File, E-mail, SGML, HTML Web Pages 등의 다양한 형태의 데이터 포맷을 지원하며 질의어와 문서 내용간의 상관관계를 자동으로 계산해 주는 Precision Rate 기능을 포함하고 있다.

구축사이트로는 대검찰청 범죄영상 수범정보 시스템 등이 있다.

**펄크럼 - Fulcrum**

펄크럼 SearchServer는 다중플랫폼을 동시에 지원하는 문서 검색 엔진으로 일반 PC 응용프로그램 뿐만 아니라 웹에서도 모든 검색이 가능하다.

SearchServer는 향상된 문서 검색 능력, Client/Server 분산 구조 지원, 표준화된 접속방법을 제공한다. 펄크럼은 오랜 기간 사용자에게 잘 알려진 SQL 쿼리 및 ODBC를 사용하여 다른 시스템들과의 호환성이 높다. SearchServer는 각각의 문서 형식과 위치를 그대로 유지하면서 텍스트를 다루기 때문에, 워드 프로세서나 데이터베이스 같은 기존 구조와 잘 조화될 수 있다.

구축사이트로는 국방부, 한양대 도서관 등이 있다.

## 데이터웨어 - BRS

SK컴퓨터 통신 등이 공급중인 미 Dataware사의 BRS서치는 데이터베이스 설계 기능에 유연성을 제공해 사용자가 필요에 따라 원문과 가까운 형태로 재구성 할 수 있는 것이 특징이며 이기종 하드웨어와 데이터베이스간의 정보 교류를 위해 Z.39.50 프로토콜을 채택하고 있다. 또한 BRS서치는 연산자와 용어의 결합 기능, 온라인 검색 중지 기능, 시소러스 기능 등을 포함하고 있으며 한글, 영문, 한자는 물론 유럽문자, 히브리어, 태국어 등 다국어어를 지원한다.

구축사이트로는 첨단학술정보센터, 체육과학연구원, 특허지원센터, 대우전자, 특허청, 에너지 연구소 등이 있다.

## 나눔기술 - 컴파스

나눔기술이 부산대와 공동 개발한

컴파스는 인터넷 탐색 도구인 로봇 에이전트와 같이 개발됐으며 기본 형태소 분석과 동시에 띄어쓰기 오류 어절 처리와 미등록어 추정기를 사용해 잘못된 분석을 최소화했다. 컴파스는 디렉토리 서비스와 웹 검색엔진이 분리되어 있는 것이 특징이다. 나눔기술은 컴파스를 그룹웨어에 옵션으로 탑재해 인터넷 시장을 집중 공략할 방침이다.

## 엔피아시스템 - 스파이더

엔피아시스템에서 개발한 스파이더는 정보검색 분야 국제 표준인 Z.39.50 프로토콜을 지원하며 한글, 한자, 숫자, 기호뿐만 아니라 영문도 자동색인해주기 때문에 해외 시장 진출도 가능한 제품으로 평가되고 있다. Z.39.50 프로토콜하에서 멀티미디어 정보 검색과 SGML, MARC까지 지원한다. 또한 전문을 대상으로 검색과 색인 작업을 수행할 뿐만 아니라 원문과 색인정보를

DBMS에 통합 관리할 수 있다.

구축사이트로는 포항공대, MBC 등이 있으며, 특히 MBC는 이를 활용한 멀티미디어 정보로 이루어진 방송 데이터베이스를 구현하고 있다.

## 연구개발정보센터 - KRISTAL-Ⅱ

연구개발정보센터에서 개발한 KRISTAL-Ⅱ는 외국으로부터 도입된 정보검색시스템의 장점을 반영하고 문제점인 한글자료 처리의 한계, 시스템 기능의 확장이나 변경의 한계, 고가의 도입비용 등을 해결하기 위해 개발되었다.

KRISTAL-Ⅱ는 문서, 영상 등 다양한 형태의 멀티미디어 정보를 관리할 수 있도록 설계되었으며 한글 자동색인을 바탕으로 한글 및 영어 문서 모두를 검색할 수 있다. 또한 웹과 윈도우즈용 인터페이스 구축을 위한 편리한 도구들을 제공한다.

〈표〉 각사의 검색엔진 비교 결과치

### 1) 검색연산관련

구분 제품명 비교내용	국내 개발						외국 제품		
	InfoACE	다센21	레이다	소프트넷	마이더스	심마니	SEARCH 97	FulCrum	BRS
개발회사	쓰리소프트	다센	한국정보공학	소프트와이즈	현인정보	한글과컴퓨터	Verity	FULCRUM	BRS
공급사	쓰리소프트	다센	한국정보공학	소프트와이즈	현인정보	데이콤	쓰리소프트	삼보정보 외	한미데이터
Boolean 검색	YES	YES	YES	YES	YES	YES	YES	YES	YES
Near 검색	YES	YES	YES	YES	YES	NO	YES	YES	YES
시소러스 검색	YES	YES	YES	NO	YES	NO	YES	YES	YES
필드별 검색	YES	YES	YES	YES	YES	NO	YES	YES	YES
날짜 조건 검색	YES	YES	YES	YES	NO	NO	YES	-	-
비교 연산자 검색	NO	YES	YES	-	NO	NO	YES	-	-
Wild Card 검색 (절단검색)	YES	YES	YES	YES	YES	NO	YES	YES	YES
자연어 검색	YES	YES	YES	YES	YES	NO	YES	YES	YES

구분 제품명 비교내용	국내 개발						외국 제품		
	InfoACE	다센21	레이다	소프트넷	마이더스	심마니	SEARCH 97	FulCrum	BRS
결과 재검색	YES	YES	NO	NO	YES	NO	YES	YES	YES
색인어 열람에 의한 검색	YES	YES	YES	YES	YES	NO	YES	YES	YES
구검색	YES	NO	NO	NO	NO	NO	YES	YES	YES
Mini-DB 검색	YES	YES	YES	YES	YES	YES	YES	YES	YES
DBMS 지원방식	DB gateway	YES	ODBC	YES	YES	NO	DB gateway	YES	ODBC

2) 색인방식

구분 제품명 비교내용	국내 개발						외국 제품		
	InfoACE	다센21	레이다	소프트넷	마이더스	심마니	SEARCH 97	FulCrum	BRS
Full-Text Indexing	YES	YES	YES	YES	YES	YES	YES	YES	YES
Remote Site Indexing	YES	YES	NO	YES	NO	YES	YES	YES	YES
한글, 영어, 한자 자동변환	YES	YES	YES	YES	YES	YES	YES	NO	YES
지원문서형식	TEXT MS-Office HWP, PDF 훈민정음 HTML SGML	TEXT WORD SGML MARC	TEXT HTML SGML Mail News	TEXT MS-WORD HWP HTML	TEXT WORD IMAGE HTML SGML	HWP DOC	TEXT WORD IMAGE HTML SGML HWP DOC	TEXT MS-Office HWP, PDF 훈민정음 HTML/SGML News	TEXT MS-Office HWP HTML Office97 Lotus Notes
색인결과와 본문압축 기능	YES	NO	NO	NO	NO	NO	NO	NO	YES
색인어 후통제	YES	YES	NO	NO	NO	NO	NO	NO	NO

3) 검색엔진별 특징

검색엔진	구조	검색기능	검색보조기능	문서지원	DB/Gateway 연동	기타	강조분야
InfoACE	형태소분석기 사전관리기 시소러스 관리기 검색엔진 DB Gateway 웹 로봇 Agent	불리언검색 절단검색 /Range 검색 가중치검 색 자연어검색 정형/비 정형 통합검색 근접검 색 시소러스검색 검색 결과 재검색	색인어 조회 가중치 Ranking 검색어 하이 라이트 시소러스조회/ 편집 검색결과정렬 및 티DB검색	Text HTML E-mail HWP, PDF SGML MS-Office 훈민정음	Oracle Informix Sybase MS-SQL Web-Robot	· 색인어 후통제 · Summarizing(주제 문장추출) · Remote Indexing · Push방식의 능동적 정보제공	· 데이터공유(운영환경 및 플랫폼에 무관) 자 체개발에 따른 강력 한 커스터마이징
다센21	형태소분석기 검색엔 진 웹 로봇	색인어검색 불리언/비 교연산자 전방/후방/중 양/양방 일치검색 필드 제한검색 사용자 히스 토리검색 열람어에 의 한 검색	시소러스검색 색인어 가중치부여 검색결과정 렬		Oracle Informix Sybase CISAM Robot	사용자관리 사전관리 보안관리 사용자로그 시 스템 프로세스 모니터 링	· 부한 경험축적에 의 한 안정화된 정보검 색시스템 제공 · 순수 국내기술로 개 발
레이다	형태소분석기 질의어 분석기 질의어확장기 시소러스관리기 검색 엔진 자원 관리기 (Lock Manager)	불리언검색 절단검색 시소러스 참조검색 필드조합검색 정형/비정형 통합검색 비교검색 색선검색	탐색어(색인어)조회 시 소러스조회 가중치기법 자연어처리 멀티 DB 검색	Text HTML News E-mail 미래로	DBMS 연동구조		정형/비정형 통합검색

이어서 →

검색엔진	구조	검색기능	검색보조기능	문서지원	DB/Gateway 연동	기타	강조분야
소프트와이즈	질의어해석기 문서변환기 소프트웨어 아이넷 넷로 씨봇	불리언검색(AND,OR)	질의어해석기 자연어검색 결과재검색 멀티 DB 검색결과정렬 하이라이트기능		Text HWP MS-Word HTML 아리랑 로터스1-2-3	· 텍스트/이미지 저장/ 검색 가능 · Smart Save 저장구조	· 자연어검색기 · 최선의 검색률 · 다양한 개발환경 · 빠른 검색속도
심마니	형태소분석기 SMDK Smrepository (DB관리)	불리언연산기능	발음유사성확장 한자어 독음 맞춤법 오류교정 한글, 영어 유의어 가중치 Ranking	HWP HTML MS-Word RTF	다양한 DB 지원	· Multi-Threading 엔 진 · 멀티서버	· 웹 환경의 편리한 유저 인터페이스 제공
MIDAS2	문서등록시스템 형태소분석기 색인어추출시스템 검색엔진 질의어해석기	불리언검색 자연어검색 질의어재검색 인접어검색 절단어검색 필드별 제한검색 결과재검색	질의어와 문서 상관관 계 계산(지능검색) Multi-Bucket 이용한 검색결과 자동정렬 Multi-DB 하이라이트 기능 이미지 Zoom 기능	Text HTML MS-WORD OCR SGML	All RDBMS	· DB별/사용자/그룹별/ 문서별 보안레벨 제공 · 사용자료명/접속시간/ 접속자 현황 등의 Account 기능 · Double-Hashing 및 Smart Save 저장구조	· 가장 빠른 색인속도 · 가장 빠른 검색속도 · 한국어 자동교정
SEARCH 97	형태소분석기 검색엔진 질의어해석기 DB Gateway Document Filter Web Robot 시소러스	불리언검색/절단검색 가중치검색 근접검색 (구/문장/단락) 자연어검색 정형/비정형 검색 시소러스 검색 사용자히스토리 검색/ 재검색 QBE검색	색인어조회 가중치 Ranking Multi-DB검색 하이라이트 기능 검색결과 정렬	Text HTML E-Mail HWP, PDF SGML MS-Office 훈민정음	Oracle Informix Sybase ODBC Web Site File System MS Exchange Lotus Notes	· Summarize · Clustering Filter · Clustering Gateway Customize · 사전관리 · Multi-Server 지원	· 다양한 플랫폼 지원 · 색인공유(플랫폼 무 관) · 강력한 customizing · 다양한 검색/안정된 검색속도
Fulcrum	Knowledge Server Knowledge Builder- Toolkit Information Server	불리언검색 Near 검색	검색결과 제한 Knowledge Map	HTML PDF MS-Office Lotus Notes	MS-Exchange server Lotus Notes Web Site File System	· Remote Document 와 색인의 동기화 · Clustering	· 다양한 종류의 문서 집합 분류 가능 자료 의 보안성 제공
BRS	형태소분석기 Robot Agent ANSI 시소러스 검색엔진 Source Manager 문서필터	논리연산(불리언검색) Wildcard검색 시소러스검색 히스토리검색 문장접근검색 단어접근검색	색인어참조기능 이미지주석기능 통합이미지지원 멀티 DB 검색	본문/색인 압축 SGML HTML Office97 Lotus Notes	ODBC Z39.50지원 4GL접속개발	· 1Terabyte 정보저장 · 사용자ID부여 · DB접근/문서/필드접 근 제한 · HyperText Link manage(Robot/Me tasearch)	· Big Data의 안전성 Gateway 확장성