

INTERNET WORLD

高祥原¹⁾

검색엔진은 보통 'Spider' 혹은 'Crawler'라 불리는 검색로봇이 인터넷을 항상 돌아다니면서 자신의 서버 컴퓨터에 색인해 놓은 데이터베이스에 기초하여 검색자가 약속된 검색형식으로 정보를 요구할 때 신속하게 문서를 검색한다. 이번 호에서는 가장 대표적인 검색엔진인 ALTAVISTA의 사용법에 대해서 간략히 알아보고, 대표적인 검색엔진들을 소개한다.

ALTAVISTA는 Digital Equipment Co.에 의해 만들어졌으며 1995년 12월에 서비스를 개시하였으며 현재 가장 선호되는 검색엔진으로 자리잡았다고 해도 과언이 아니다. DEC의 주장에 의하면 이 검색엔진은 현재 47만 6천개의 서버에 의한 3천 1백만 문서에 접근할 수 있도록 해주며, 하루 2천 6백만건의 조회를 처리하고 있다. ALTAVISTA의 주소는 <http://www.altavista.digital.dom>이다. 이 검색엔진을 이용하여 문서를 검색하는 방법은 Simple Query와 Advanced Query의 두가지로 나누어지는데 아래서는 각각에 대해 간단히 설명하기로 한다.

Simple Query

우선 ALTAVISTA는 대문자와 소문자를 구분한다. 예를 들어 터키(국가)에 대한 정보를 얻고 싶을 때 'Turkey'로 입력하면 소문자로 칠면조(turkey)를 나타낸 문서를 제외시킬 수 있다. 이 검색엔진은 절단기능도 지원하는데, 검색하고자 하는 단어를 확실히 모르는 경우에 이 기능을 사용하면 매우 편리하다. 예를 들어 '성공하는 사람들의 7가지 비결'을 지은 저자의 성은 Covey인 것을 아는데 이름이 Steve인지 Steven인지 혹은 Steplen인지를 모를 때 검색어에 Ste*Covey를 입력하면 된다. *를 사용할 때 주의할 점은 반드시 *앞에 영문 3문자가 포함되어야 하며, *내에 포함될 수 있는 영문자는 최대한 5자라는 점이다.

검색식에서 '+'를 사용한 것은 '+'뒤에 나오는 단어가 반드시 포함되어야 한다는 의미이다. 단어를 계속 나열하면 연산자 OR과 같은 결과를 출력하고, '-'부호를 사용함으로써 그 뒤에 나오는 단어는 반드시 나타나지 않아야 되도록 검색을 제약할 수 있다.

문서의 특정부문만을 검색하는 제한검색기능도 가능한데, 제한검색을 하기 위해서는 '검색을 실시할 필드명: 검색어의 형식'으로 입력한다. 예를 들어 Kist와 링크되어 있는 문서들을 찾고 싶다면 link:kist.re.kr을 입력하면 된다. 제한 검색을 할 수 있는 필드는 anchor(hyper link에 포함된 단어검색), applet(자바 애플릿 검색), url(페이지의 url 검색), title(제목검색), text(본문 검색), link(연결주소 검색), image(화상검색)등이 있다. 검색 결과를 출력하는 방식은 standard, compact, detailed의 3가지 유형이 있는데, standard는 문서제목, 문서의 첫부분 150문자, URL, 문서의 크기, 등록날짜를 출력하고, compact는 제목 25문자, 문서의 첫부분 30문자, 등록날짜를 출력하며, detailed는 standard와 똑같은 내용을 출력한다. 검색을 한 후에 검색 결과의 문서내에서 검색어로 입력한 단어를 찾으려면 Netscape의 edit행에서 view를 실행하면 된다.

Advanced Query

이 검색방법은 다양한 연산자를 사용할 수 있고, 출력에 있어 우선순위 단어를 선택할 수 있으므로 보다 정밀한 검색을 하는 데 적합하다. Simple Query의 대소문자 구분기능, 절단기능, 제한검색기능 등은 모두 가지고 있다. 그러나 구를 검색할 때는 " "를 사용한다. Advanced Query는 검색어를 입력하는 Selection Criteria란, 출력 우선순위단어를 입력하는 Results Ranking Criteria란 문서가 등록된 기간의 범위를 설정하는 date란으로 구성되어

있다. Advanced Query에서 사용할 수 있는 연산자는 AND(&), OR(|), NOT(!), NEAR(~)이다. &는 연산자 좌우에 입력된 단어가 동시에 포함한 문서만을 검색하라는 명령이고, |는 연산자 좌우에 입력된 단어를 하나라도 포함하고 있는 문서를 검색하라는 명령이다. !는 이 연산자 뒤에 나오는 단어를 포함한 문서는 검색결과에서 제외시키라는 명령어이며, ~는 연산자 좌우에 입력된 순서에 관계없이 두단어가 10단어 이내에 함께 나타난 구를 포함하고 있는 문서들을 모두 출력하게 된다. 이들 연산자들의 우선 순위는 NEAR(~)→NOT(!)→AND(&)→OR(|)이다. 검색출력방식은 Simple Query와 동일한 3가지 방식 이외에 검색된 건수만을 출력할 수 있는 Count Only가 추가로 포함되어 있다.

ALTAVISTA 외에도 다양한 검색엔진들이 있는데 이들의 주소와 대표적인 특징은 다음과 같다.

Excite(<http://www.excite.com>): Concept Search를 제공함. 즉 어떠한 검색어를 입력하면 이 검색어와 관련된 개념을 포함한 모든 문서를 찾아줌. 평범한 문장을 검색어로 사용할 수 있음.

Galaxy(<http://www.einet.net/galaxy.html>): 제한검색, gopher, telnet검색이 가능함.

Infoseek Ultra(<http://ultra.infoseek.com>): 이 회사의 주장에 의하면 가장 빠르고 방대한 자료를 가지고 있는 차세대 검색엔진임. 실시간 색인을 'Super Spider'란 로봇이 실행하기 때문에 가장 update되어 있다고 주장됨.

Inktomi(<http://inktomi.berkeley.edu>): 기호검색이 가능함.

Lycos(<http://www.lycos.com>): Top 5% Sites, Point Reivew를 포함하고 있음.

Magellan(<http://www.mckinley.com>): 보유문서 전체에 대한 검색을 할 수도 있고 우수한 사이트를 모아놓은 Rated and Reviewed 데이터 베이스 내의 자료만을 검색 할 수도 있음.

WebCrawler(<http://webcrawler.com>): American Online사가 운영하는 검색엔진. 평범한 문장을 검색어로 사용할 수 있음.

Yahoo!(<http://www.yahoo.com>): 로봇이 아닌 사람이 정보를 수집, 가공, 분류함. 양적인 한계를 극복하기 위해 ALTAVISTA검색결과도 함께 보여줌. Weekly Picks, Daily Picks를 통해 우수한 사이트를 선정함.

한국의 대표적인 검색엔진으로는 다음과 같은 것이 있다.

심마니(<http://simmany.hnc.net>): 한글과 컴퓨터가 개발한 검색엔진. 절단기능을 가지고 있으며, 연산자도 사용할 수 있음.

코시크(<http://kor-seek.chungnam.ac.kr/cgi-bin/korea>): 충남대학교의 검색엔진

정보탐정(<http://aistar.kotel.co.kr>): 한국통신의 검색엔진. 가장 속도가 빠르다고 주장됨. 절단 기능은 %로 표시함.

한국의 디렉토리(<http://www.dir.co.kr/>): 검색엔진은 아니지만, 우리나라의 주요 사이트를 주제별로 잘 분류해 놓고 있다. 신문기사를 찾는다는 각 신문사의 홈페이지를 방문하면 대부분의 홈페이지가 기사검색기능을 가지고 있다. 아래는 한국의 디렉토리가 모아놓은 신문

사의 주소이다.

경향신문[<http://www.khan.co.kr/news/headliness/index.htm>]

내외통신[<http://zec.three.co.kr/koreaherald/nknews/nkindex.html>]

디지털 조선일보[<http://www.chosun.com>]

매일신문[<http://www.m2000.co.kr>]

부산일보피스넷[<http://www.puil.com>]

서울경제신문[<http://www.korealink.co.kr/sknew/seoul.htm>,]

서울신문 [<http://www.seoul.co.kr/seoul/seoul.htm>]

서울신문 뉴스넷[<http://www.seoul.co.kr>]

스포츠 서울[<http://www.seoul.co.kr/sports/sport.htm>]

일간스포츠[<http://www.korealink.co.kr/sport/sport.htm>]

중앙일보[<http://www.joongang.co.kr>]

코리아 네트워크[<http://tradenews.co.kr/korean/korean.html>]

한겨레 신문[<http://news.hani.co.kr/hkr>]

한겨레 Multimedia World[<http://news.hani.co.kr>]

한국경제신문[<http://www.ked.co.kr>]

한국일보[<http://www.korealink.co.kr/hankook/hankook.htm>]

현대불교신문사[<http://www.buddhapia.co.kr/hyundai/hyun01.html>]

DAILY TRADE NEWS of KOREA[<http://tradenews.co.kr>]

kinds 신문기사 검색(한국언론연구원)[<http://www.kpi.or.kr>]

Korea Herald [<http://zec.three.co.kr/koreaherald>]

KoreaTimes[<http://www.korealink.co.kr/times/times.htm>]

이 중 특히 한국언론연구원의 Kinds(Korean Integrated Newspapers Database System)는 신문기사검색을 하는데 매우 유용하다.

주석 1) 혁신체제팀 선임연구원, 경제학 박사(Tel: 01-250-3034)

