

국내 웹검색도구의 특성 및 탐색 기능 평가에 관한 연구

A Study on the Performances of Korean WWW Search Tools

이 란 주(Lan-Ju Lee) *

최 경 화(Kyung-Hwa Choi) **

목 차

- | | |
|-------------------------|----------------------------|
| 1. 서론 | 3. 2. 1. 2 메타검색엔진 |
| 2. 연구방법 | 3. 2. 1. 3 통합검색엔진 |
| 3. 검색도구 | 3. 2. 2 검색엔진의 정보검색법 |
| 3. 1. 목록서비스 | 3. 2. 2. 1 각 검색엔진의 개별분석 |
| 3. 1. 1 목록서비스의 유형 | 3. 2. 2. 2 각 검색엔진의 정보검색 옵션 |
| 3. 2. 1 목록서비스의 비교분석 | 3. 2. 2. 3 검색엔진의 정보검색 옵션 |
| 3. 2 검색엔진 | 비교분석 |
| 3. 2. 1 검색엔진의 유형 | 3. 3 검색도구의 활용방안 |
| 3. 2. 1. 1 단어별 검색엔진의 유형 | 4. 결론 |

초 록

최근에 국내외의 웹문서들을 대상으로 한글의 특징적인 면을 고려해서 편리하게 이용할 수 있는 국내 검색도구들이 개발되고 있다. 본 연구의 목적은 이러한 국내 검색도구들의 종류를 살핀 뒤, 이들의 특성을 파악해 이용자 입장에서 정보요구에 적합한 검색도구를 선택하도록 돕는데 있다.

검색도구를 크게 목록서비스(코시크, DIR, ZIP, 넷퀘스트21, 심마니)와 검색엔진(심마니, 까치네, 정보탐정, 애니서치, 유니파인더, 웹글라이더, 미스다찾니)으로 나누고 이들의 현황, 검색옵션과 장단점 및 성격등을 비교·평가함으로써 정보요구에 따른 검색도구 선택에 대한 방법론을 제공한다.

또한 국내 검색도구의 문제점을 지적함으로써 보다 향상된 검색도구의 설계에 도움이 되고자 한다.

ABSTRACT

The purpose of the article is to help users to select appropriate Korean WWW search tools and to retrieve Web documents for their information needs by using them effectively. It analyzes the characteristics, functions, and advantages and disadvantages of each search tool, while search tools are divided broadly by two categories - subject directory services (Kor-seeK, DIR, ZIP, Netquest21, Simmany) and keyword search engines (Simmany, Kachine, Jungbotamjung, Anysearch, Unifinder, Webglider, Missdachanni). It provides the selection criteria of search tools for users who have information needs. In addition, the study also attempts to contribute to the improvements in designing Korean WWW search tools, as investigating their shortcomings.

* 동덕여자대학교 문헌정보학과 조교수

** 동덕여자대학교 대학원 문헌정보학과

접수일자 1997년 8월 14일

1. 서론

현대사회가 정보화 사회로 나아감에 따라 좀더 많은 정보를 빠르게 얻기 위한 방법들이 개발되어 사용되고 있다. 그 중에서 가장 많은 정보를 가장 신속하게 구할 수 있는 방법은 바로 인터넷을 이용하는 것이다. 인터넷은 실로 방대한 정보를 전세계 각지에 가지고 있으면서 필요로 하는 정보와 지식을 제공해 줄 수 있는 유용한 자원이다. 그러나 인터넷은 전세계에 걸쳐 연결된 방대한 규모이므로 인터넷에서 자신이 필요로 하는 정보를 찾아낸다는 것은 그리 쉬운 일이 아니다. 최근 들어 이용자들이 원하는 정보를 간편하게 찾을 수 있도록 도와주는 각종 검색도구가 개발되고 있다. Alta Vista, Lycos, Infoseek, Excite 등은 국외의 유명 검색도구들로서 인터넷을 향해하는 국내 이용자들이 많이 이용해 오고 있다. 그러나 이러한 국외의 검색도구들은 영문으로 된 자료가 적은 국내 사이트들의 문서에 대한 배려가 적고 대부분의 국외 검색도구들은 한글로 된 검색단어를 이용할 수도 없다. 따라서 최근 몇 년 동안에 국내외의 웹문서들을 대상으로 검색할 수 있을 뿐만 아니라 한글의 고유한 특성을 이용해서 편리하게 검색해 주는 검색도구들이 개발되고 있다. 데이터베이스의 양적인 면에서 국외 검색엔진의 로봇보다 훨씬 작고 아직은 많은 검색도구들의 검색 대상이 국내 웹으로 한정되어 있기 때문에 다양한 정보 습득이 어렵지만 기능들이 점차 향

상되고 있으니 기대해 볼 만 하다. 또한 이러한 국내 검색도구들의 개발은 일반 이용자와 웹사이트 운영자에게도 모두 바람직한 것이라고 할 수 있다. 국내 검색도구들은 특히 인터넷에 익숙하지 못한 이용자에게 편리한 인터페이스를 주고 있기에 이에 대한 연구는 의의가 있다고 하겠다. 본 연구의 목적은 이용자 입장에서의 그들의 정보요구에 적합한 국내 검색도구의 선택과 효율적인 사용을 돕고자 하는 것이다. 따라서 국내 검색도구들의 종류를 살핀 뒤, 각 검색도구들의 특성 및 탐색기능을 조사 분석한다.

2. 연구방법

최근 들어 이러한 웹에서 이용자들이 필요로 하는 정보를 쉽게 찾을 수 있도록 도와주는 각종 검색도구들과 색인들이 계속적으로 개발되고 있으며, 이들 검색도구의 성능을 평가하는 연구 또한 진행 중에 있다. 그리고 이러한 연구들은 각 검색도구들이 제공하는 탐색기능이나 데이터베이스의 구성 방법들을 성능 평가의 기준으로 하여 비교하는 연구(Sullivan, 1997 ; 이란주, 1997)와 이들에 대해서는 간단히 언급하고 탐색질문을 이용하여 검색한 뒤 검색결과를 가지고 각각의 검색도구들을 비교하는 연구가 이뤄지고 있다(Chu, H. and Rosenthal, M., 1996 ; Courtois, et al., 1995 ; Kimmel, 1996 ; Leighton, 1995 ; Leonard, 1996 ;

Scoville, 1996 ; Zorn et al., 1996, ; 정영미 김성은, 1997). 각 연구들의 결과는 방법론과 평가기준이 다르고 선정된 검색 도구들도 다르기 때문에 서로 다른 결과를 보이고 있다. 그러나 만약 검색도구들을 평가하는 목적이 이용자로 하여금 원하는 정보를 얻기 위해 검색도구를 사용할 때, 가장 자신에게 적합한 검색도구가 무엇인가를 판단하는데 기여하도록 하는데 그 목적이 있다면, 각각의 검색도구들의 특성을 파악해서 각기 다른 정보요구에 맞는 검색도구들을 찾을 수 있도록 안내하고 그 검색도구를 사용해서 가장 높은 효율을 낼 수 있는 검색방법들을 알려 줄 수 있는 평가방법들이 필요하다고 보여진다.

본 연구에서는 우선 검색도구의 유형을 목록서비스와 검색엔진으로 나누고 검색엔진은 다시 단어별 검색엔진과 메타검색엔진 그리고 통합검색엔진으로 나눈다. 다음 단계는 각 유형에 속하는 국내 목록서비스(코식크, DIR, ZIP, 넷퀘스트21, 심마니)와 검색엔진(심마니, 까치네, 정보탐정, 웹글라이더, 애니서치, 유니파인더, 미스다찾니)을 그들의 성격, 특성, 기능별로 조사 분석한다. 특히 목록서비스의 경우에는 현재 국내에서 운영되고 있는 “웹진”이 전부 몇 개인지를 알아보고 이것을 통하여 이용자 입장에서의 국내 목록서비스의 주제 카테고리상의 문제점을 조사한다. 검색엔진의 경우에는 검색식을 작성하는데 사용되는 정보검색법에 대해 각 검색엔진들이 갖는 검색옵션을 비교 분석한다. 그

리고 마지막으로, 분석된 국내 검색도구들의 현황과 특성 및 탐색 기능을 비교 평가해서 검색도구의 설계자에게는 이상적인 검색도구의 모형을 제시하고 이용자에게는 정보요구에 따른 최적의 검색도구 선택에 대한 하나의 방법을 제시하고자 한다. 본 논문의 연구방법으로는 문헌조사와 각 검색도구들의 도움말 기능과 실제 검색도구들을 사용해 본 결과를 종합한 것이다. 이 연구의 제한점은 웹정보자원이 갖는 유동성과 더불어 검색도구들도 새로운 모습으로 자주 추가 삭제되기 때문에 본 연구에서 분석 평가된 결과들이 계속 유지되지 않는다는 것이다. 따라서 검색도구들에 대한 지속적이며 다각적인 연구의 필요성이 여기에 있다고 보여진다. 또한 지면 관계상 검색도구의 특성을 좀 더 잘 보여 줄 수 있는 그림을 포함하지 못한 제한점을 갖고 있다.

3. 검색도구

검색도구의 유형은 이용자 입장에서 살펴볼 때, 첫째 목록서비스와 둘째 검색엔진으로 나눌 수 있고 검색엔진은 다시 단어별 검색엔진과 메타검색엔진 그리고 통합검색엔진으로 나뉘어진다. 본 장에서는 각 유형에 대한 정의와 검색도구들의 종류 및 특성을 살피고자 한다.

3.1 목록서비스

목록서비스란 특정 주제별로 각 페이지들을 분류하여 정리해 놓은 것으로써 웹 사이트들의 주제별 목록(directory)을 제공하고 있다. 즉 수많은 사이트들을 찾아 보기 쉽도록 유사한 주제의 정보를 제공하는 곳끼리 분류가 되어 있고, 원하는 분야를 마우스로 클릭하면 그 분야의 세부 항목이 나온다. 이러한 방식은 도서관에서 많이 사용되는 10진분류법과 유사한 방법으로서 오래전부터 이용되어 온 매우 자연스러운 검색방법이다. 이런 식으로 사용자가 계속 링크해서 들어가다 보면 본인이 원하는 정보를 찾을 수가 있다.

이러한 서비스 방식은 다양한 종류의 사이트가 많이 만들어지면서 매우 많은 수가 운영되고 있지만 그중에서 널리 알려진 곳은 DIR, 코식크, 넷퀘스트21, ZIP, 심마니, Korea Directory, WWW Servers in Korea(동국대학교), WWW Servers in Korea(충남대학교), Active Map for WWW Servers in Korea(포항공과대학교) 등이다.

목록서비스의 핵심은 “주제분류”이므로 효과적으로 목록서비스를 이용하기 위해서는 각 목록서비스의 부주제분류를 잘 이해하고 있어야 한다. 그렇지 않으면 한번 길을 잘못 들어서면 본인이 원하는 정보가 아닌 다른 정보를 찾게 되는 경우가 생길 수 있다.

이 서비스의 장점은 찾고자 하는 것에 대하여 아무런 지식이 없어도 원하는 것의 대분류 정도만 알아도 찾을 수 있다는 것이지만 원하는 정보를 얻기까지 여러단

계를 거쳐야 하므로 중간에 길을 잘못 들어서면 원하지 않는 정보를 찾게 될 수도 있는 단점을 가지고 있다. 그런데 위에 나열된 사이트들은 국내 사이트의 목록을 제공하고 있다는 점은 같지만 모든 곳에서 정보검색에 적합한 주제별 목록을 제공하고 있지는 않다. Korea Directory는 국내 사이트의 부분적인 주제별 목록과 일반 텍스트 정보, 해외 사이트 목록등 복합적인 형태의 정보를 제공하고 있다. WWW Servers in Korea(동국대학교)와 Active Map for WWW in Korea(포항공과대학교)는 World Wide Web Consortium(W3C)에 공식 대한민국 사이트 목록으로 등록되어 있는 것이지만 주제별 목록을 제공하지 않고 오직 사이트 소재지별 목록을 제공하고 있다. 또한 WWW Servers in Korea(충남대학교)는 두번째 단계 도메인(co.kr, ac.kr 등)별로 사이트의 목록을 제공하고 있다. 이들은 주제별 목록을 제공하지 않기 때문에 일반적인 정보검색을 위한 목록서비스로 보기 어렵다. 그러므로 이 연구에서는 정보검색에 적합하도록 주제별 목록을 제공하는 DIR, 코식크(Kor-seek), 넷퀘스트21(NetQuest21), ZIP, 심마니에 대해서만 분석하고자 한다.

3. 1. 1 목록서비스의 유형

1) 코식크

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

→ 국내 최초의 검색엔진

국내 최초의 검색도구인 코식크는 충남 대의 한아름이 제작해 제공하는 사이트로서 현재는 웹코리아로부터 지원을 받고 있다. 한글과 영문으로도 서비스하고 있으며 주제어에 따른 검색기능과 각 분야별 검색, 그리고 여론광장, 새로운 웹, 웹등록, 국내외 사이트 검색, 웹만들기 등의 기능별 서비스 제공으로 구성되어 있다. 특히 여론광장은 일종의 BBS 형태로 운영하고 있어, 웹상에서의 전자게시판의 역할을 제공해 준다. 자체에 데이터베이스 검색기능을 가지고 있으며 필요없다고 판단되는 부분을 빼버렸기 때문에 빠르게 분류 이동을 할 수 있는 장점을 가지고 있다. 그러나 SUN SPARC classic, IBM-PD 386SX-21이라는 개발환경때문인지 검색속도가 느린편이다. 또한 사용자들을 위한 도움말 기능이나 자세한 설명이 없어 제목만 보고 이동해야 하는 불편이 있다.

코식크 검색방법의 단점중의 하나는 연산자가 없어 복잡한 검색식은 사용할 수 없다는 것이다. 대체로 원하는 정보를 찾기 위해서는 보편적인 단어를 다수 사용해야 하는데 다른 검색엔진들에 비해 성능이 떨어지는 편이므로 주제분류검색시 사용하는 것이 좋다고 본다. 분류검색, 주제검색, 확장검색을 지원하며 분류검색을 통해 영문과 한글로 제공되는 분야별 웹 사이트를 검색할 수 있다.

(1) 분류검색

총 11개의 디렉토리로 되어 있는데 한글 디렉토리 서비스를 통해 오랫동안 데

이터베이스를 축적하여 관련 분야별 정보가 체계적으로 정리되어 있다.

(2) 주제어 검색

입력된 단어는 쉼표나 공백(뛰어쓰기)으로 구분되며 각 단어들은 OR로 연산되어 검색어에 해당되는 단어들이 모두 포함된 문서에서 시작하여 마지막 한 단어까지 포함된 문서를 순서적으로 출력해준다. 홈페이지와 주제별 선두문서를 위주로 검색을 수행함으로써 원하는 정보를 찾기가 쉽지 않은데 이런 때는 "확장검색"을 사용하는 것이 좋다.

2) DIR

(<http://www.dir.co.kr>)

→ 한국의 디렉토리 서비스

DIR은 국내 뿐만 아니라 외국에도 상당히 알려져 있다. 외국의 검색엔진인 익사이트에도 등록되어 있으며 국내 디렉토리 서비스 사이트 중에서 가장 앞서가고 있다고 할 수 있다. 국내 웹 사이트 목록 서비스와 웹 관련 뉴스 게시판 서비스의 두 가지 서비스를 제공하고 있다. 여기서는 12개의 주요 주제분류 및 상세한 세부 주제분류를 이용해서 국내 사이트들을 분류한 후에 소개하기 때문에 특정한 주제의 내용을 다루는 사이트를 찾기가 용이하다. 그리고 무엇보다도 각 사이트에 대해 각 사이트의 운영자가 보내온 소개 자료를 제공하고 있기 때문에 각 사이트에 대한 보다 정확한 이해가 가능한 잇점을 가지고 있다. 사이트 정보표시 영역을 자세히 살펴보면 사이트제목, 사이트 주소, 운영

자, 주제분류, 운영목적, 컴퓨터의 소재지, 운영개시일, 사용언어, 갱신주기, 회선속도, 주요내용, 사용기술, 관련뉴스, 서버사이트들이 있다. 대부분의 검색엔진들이 가중치 순에 따라 검색결과를 보여주지만 검색한 정보에 대한 자세한 설명이 없어 검색해 낸 정보들의 사이트로 접속해서 확인하는데만 해도 상당히 많은 시간이 걸린다고 볼 때, DIR의 이러한 자세한 정보영역표시는 대단히 유용하다고 볼 수 있다. 뿐만 아니라 해외의 사이트들은 대부분 하나의 사이트에서 하나의 주제만을 다루지만 국내 사이트는 여러 개의 주제를 같이 다루는 곳이 많다. 이러한 면에서 볼 때 이곳에서의 '서브 사이트'라는 개념은 훨씬 더 유용하다. '서브 사이트'란 하나의 사이트에서 여러 개의 주제를 다루고 있다면 각각의 주제를 다루는 부분을 마치 다른 사이트인 것처럼 취급해서 별도로 분류해서 소개하는 것이다. 그러므로 주제분류를 이용해서 사이트를 찾을 때 다른 서비스보다 더 정확하게 연관된 사이트를 찾을 수 있다. 게다가 사용이 편리하고 디자인도 깔끔하다.

3) ZIP (<http://zip.org>)

→ 한국판 야후

ZIP 역시 주제별로 분류되어 있는 웹디렉토리로서 인터넷 포럼인 나우누리 인터넷 스터디 포럼의 시삽이 개발한 것으로 한국판 야후라고 할 수 있다(한국경제신문 97년 3월 1일, 문화일보 97년 3월 4일). ZIP은 1997년 2월 20일 Starting Point의

Computing 분야 Hot Site로 선정된 데에 이어서 4월 14일에는 세계 최대 웹 프로모션 서비스인 Submit it (<http://www.submit-it.com>)으로부터 한국에 선 유일하게 등록대상 사이트로 선정되었고 Looksmart International사의 편집자가 선택한 사이트로 선정되기도 했다. Starting Point는 사용자에게 의해 직접 투표로 선정되는 것으로, 웹 검색시 필수적인 도구로 인터넷 향개의 시작점인 서비스로 각 분야별 한 개씩의 Hot Site를 선별하여 전시하고 있으며 메타검색엔진 등의 서비스를 제공하고 있다(<http://www.stpt.com>). 그리고 Submit it은 각 기업과 개인의 URL을 전세계의 유명한 검색엔진과 디렉토리에 등록할 수 있도록 도와주는 웹서비스이다. 즉 사용자가 등록을 요청한 사이트들 중 300개를 엄선하여 100개의 검색엔진과 디렉토리에 등록을 대행해 준다(한글판 PC WEEK, 97년 3월호).

ZIP이 Submit it의 공식 등록사이트로 선정됨에 따라 현재 한글과 영문을 지원 하는 주제별 검색서비스 이용시 더욱 다양한 양질의 사이트를 제공받을 수 있을 것으로 보이며 국내 뿐만 아니라 전세계적으로 진출할 수 있는 교두보를 마련한 셈이 된다. 등록된 영문자료 중 추천할 만한 좋은 내용의 등록 사이트는 한글로도 번역하여 제공하고 한글 페이지에서도 해당 분류와 하위 분류의 영문자료가 한글 자료의 수와 얼마나 차이가 있는지 쉽게 파악할 수 있도록 해 영문자료의 검색이

용을 늘리도록 할 방침이다. ZIP의 가장 큰 장점은 국내 새로 등록되는 사이트들 중 나우누리 ISF내의 Web Rating Team에 의해 정리 평가된 뒤 등록이 이루어 진다는 점으로 사업자가 아닌 이용자 관점에서 이용자를 위해 사이트를 선정한다는 것이다. 현재 100여개의 주제별로 1200여개의 국내 사이트들이 연결되어 있는 ZIP은 자체개발한 서치엔진을 탑재했고, 광고엔진도 개발해 앞으로 광고서비스를 할 계획이다.

주제별 디렉토리이지만 ZIP에 등록된 사이트들을 이름, 설명, 평가, URL에 대해서도 검색어로 검색이 가능하다. 특별한 연산자는 사용하지 않으며, 두 단어 이상을 넣으면 OR로 처리하여 모두 검색해낸다. 검색결과가 나오지 않거나 다른 검색엔진을 사용하고 싶을 때는 검색결과 아래에 다른 유명한 검색엔진을 같은 찾기 명령어로 클릭하여 검색할 수 있도록 하고 있다.

4) 넷퀘스트21

(<http://www.netquest21.com>)

→ 국내 유일의 정기적인 정보사냥 대회 개최

넷퀘스트21은 서버구축, 호스팅서버 대여 및 웹페이지 구축을 전문으로 하는 인터넷 전문 컨설팅업체에 의해 만들어진 것으로 1996년 4월 22일 "Click" 이라는 이름의 웹잡지를 시작으로 이루어진 것이다. 국내 웹디렉토리 서비스를 제공하는 Inside Korea 이외에도 격주마다 개최되는

수준높은 정보사냥대회인 Web Hunting, 매일매일 가볼만한 해외 웹사이트를 선정하는 Internet Book, 최신뉴스와 소식들을 전해드리는 News Quest, 이틀마다 업데이트되는 최신 웨어웨어소식 Downloads, 사용자와의 긴밀한 대화 창구인 게시판이 웹채팅과 같은 서비스를 하고 있다. Inside Korea는 전체 10개의 분야로 되어 있으며 자체 검색엔진으로 원하는 사이트를 찾을 수 있고 웹사이트를 등록할 수도 있다. 매주 추천되는 사이트(Pick of Week)가 볼 만하다. 인사이드 코리아 역시 주제별 디렉토리이지만 검색어도 사용할 수 있도록 되어 있다. 그러나 검색어 사용에 관한 도움말이 없어 입력된 검색어를 가지고 어떤 식으로 찾아주는 것인지 정확히 알 수는 없으나 검색결과 출력된 요약정보(brief info.)내의 검색어 출현으로 볼 때 공백은 OR로 처리하고 공백이 없는 것은 AND로 처리하는 것 같다. 그러나 단어가 길어지면 역시 OR로 처리하는 것으로 보인다.

5) 심마니 (<http://simmany.hnc.net>)

→ 뛰어난 자연어 처리

마지막으로 심마니는 화면 중간에 총 16개 분야에 이르는 주제를 설정하여 해당하는 사이트를 분류해 놓았다. 그리고 각 사이트에 대한 간결한 소개가 있어 쉽게 각 사이트의 특성을 파악할 수 있다. 주제별 분류는 검색기능을 보완해 주는 방법으로 비슷한 주제를 가지고 있는 사이트를 한 눈에 볼 수 있는 편리함을 가

지고 있으며 분류된 사이트는 주로 국내 사이트가 주류를 이룬다. 대분류, 중분류, 소분류로 계층적인 분류에 의해 사이트가 나누어지며 괄호 안의 숫자는 해당 주제에 포함되어 있는 사이트의 수를 말한다. 심마니의 분류서비스는 부가적인 기능이므로 그밖의 자세한 내용은 단어별 검색 엔진에서 다루기로 한다.

3. 2. 1 목록서비스의 비교분석

지금까지 살펴 본 목록서비스는 대체로 별도의 검색옵션을 제공하지 않고 주제분류 서비스를 통해서만 원하는 정보로 접근할 수 있기 때문에 이들에 대해 좀 더 자세한 평가를 하기 위해서 위의 서비스를 이용해서 현재 국내에서 운영되고 있는 “웹진”(인터넷으로만 운영되는 잡지)이 전부 몇 개인지를 살펴 보았다. 심마니는 단어별 검색서비스를 지원하므로 단어별 검색엔진에서 검색옵션을 별도로 평가하며 코식크는 단어별 검색엔진으로의 역할을 제대로 수행하지 못하므로 목록서비스에 포함시켰다.

먼저 총 11개의 최상위주제로 분류되는 코식크는 ‘언론 종교 역사’의 대주제에서 ‘잡지’의 순으로 주제를 선택해야 된다. 출판잡지들과 웹진이 함께 나타나므로 정확한 웹진의 수를 알기 어렵다. 뿐만 아니라 사이트에 대한 이름만 제시하고 간단한 소개조차 없기 때문에 실제로 접속해 보기 전에는 웹진의 정확한 수를 알기 어려울 뿐만 아니라 출판잡지와 웹진을 구분하기조차 어렵다. 그러나 영문서비스를

지원하므로 영문으로 된 주제분류와 사이트 이름을 보여주기도 한다. 다음으로 DIR은 총 12개의 최상위 주제목록으로 구성되어 있는데 ‘언론’을 선택하면 언론에 해당하는 세부주제 목록과 언론에 해당하는 사이트의 목록이 나타난다. 이 때 세부 주제목록 가운데 ‘전자잡지’를 선택하면 사이트의 이름과 간단한 소개가 보여지는데 해당 사이트에 관한 자세한 정보를 얻을 수도 있다.

비교적 최근에 개설된 ZIP은 총 14개의 최상위의 주제분류로 되어 있는데 ‘뉴스와 미디어’라는 대주제를 선택하고 ‘잡지’라는 서브 주제를 선택하면 아주 많은 세부주제 목록을 주제별로 일반잡지와 웹진의 구별없이 보여 준다. 따라서 각 세부주제 목록을 직접 살피고 접속해야만 웹진의 여부를 알 수 있다. 출력결과는 사이트에 대한 설명과 평가를 한글과 영문으로 지원한다. 다음으로 넷퀘스트21 중 인사이트 코리아는 총 10개의 최상위 주제목록을 제공하고 있으며 ‘뉴스 언론’이라는 대주제 다음으로 ‘웹매거진’을 선택하면 웹진의 목록을 사이트에 대한 설명과 넷퀘스트21에서 덧붙인 평가를 곁들여 볼 수 있다.

마지막으로 심마니의 경우는 총 16개의 최상위 주제목록을 가지고 있는데 “웹진” 사이트의 목록을 찾기 위해서는 ‘뉴스 언론매체’라는 대주제 다음으로 ‘잡지’를 선택해야 한다. 그런데 이곳에는 일반잡지와 웹진이 함께 있으므로 실제로 사이트에 접속해 보기 전까지는 “웹진”의 정확

한 수를 알기가 어렵다. 검색된 사이트 목록에서는 각 사이트에 대한 간단한 소개를 하고 있기 때문에 “웹진”을 이해하는데 도움이 된다.

“웹진”을 검색해 본 결과 각 목록서비스마다 웹진이 있는 주제분류와 검색의 결과가 서로 다르게 나온다는 것을 알 수 있었다. 웹사이트의 성격은 아주 다양하고 풍부하기 때문에 모든 사이트들의 주제를 한 화면에 나타내게 되면 너무나 많은 주제목록이 나타나게 되므로 원하는 정보를 찾을 수가 없다. 따라서 정보에 대한 접근을 용이하게 하려면 계층구조적인 분류체계를 만들어야 하는데 이는 목록서비스를 운영하는 자가 자신의 관점에 따라 가장 적합할 것이라고 생각하는대로 즉 어떤 사이트의 주제가 무엇일 것이라고 판단하는 대로 만들기 때문에 서로 다른 분류체계를 갖게 된다. 따라서 이용자들은 분류 목록서비스를 이용할 때마다 서로다른 분류체계 때문에 혼란을 겪게 된다. 국내 단어별 검색엔진 가운데 목록서비스를 실시하고 있는 “애니서치”는 유일하게 주제분류에 문헌정보학 전문가들이 참여했는데, 이렇게 주제분류를 지원하는 모든 목록서비스의 분류체계에 전문가들인 문헌정보학 사람들이 참여했다면 지금보다 표준적이고 체계적인 주제분류들이 가능했었으리라고 본다. 목록서비스를 이용하는 정보 검색은 편리하지만 목록서비스는 결국 특정한 주제에 관한 정보를 제공하는 사이트를 검색하는 방법이라고 말할 수 있겠다.

한편 국내 인터넷의 현실은 국내 목록서비스의 서비스 수준이 외국의 경우에 비해 제약을 받고 있다. 즉 인터넷 비용이 저렴한 미국에서는 하나의 사이트에서 하나의 주제와 관련된 정보를 제공하는 것이 일반적이지만 인터넷 관련비용 즉 컴퓨터 가격, 네트워크 장비 가격, 네트워크 회선 유지 비용, 웹사이트 구축비용 등이 비싼 국내에서는 하나의 사이트에서 다양한 종류의 정보를 제공하는 것이 보편화되어 있기 때문이다. 따라서 하나의 큰 사이트 밑에 독립적인 주제를 가지는 다수의 서브들이 있는 것이다. 즉 사이트는 하나일지라도 해당하는 목록서비스의 주제분류는 여러 개가 되는 경우가 많은 것이다. 그러나 사이트의 운영자가 목록서비스에 등록하거나 목록서비스 운영자가 등록하는 경우에 그 사이트에서 제공하는 모든 정보에 적합한 주제분류로 모두 등록시키지 않고 가장 대표적인 정보에 해당하는 주제분류로만 분류하는 경우가 많기 때문에 상당히 많은 양의 정보가 사장되고 있다. 주제분류에 지정되지 않은 정보는 목록서비스를 이용해서 검색할 수 없기 때문이다(W3, 6월호).

DIR과 같은 일부 목록서비스에서는 그와 같은 서브사이트에 해당하는 주제분류로도 분류하고 있기는 하지만, 이러한 문제는 궁극적으로 각 사이트의 운영자가 자신이 운영하는 사이트에서 제공하는 정보를 정확히 파악하고 적극적으로 목록 사이트에 등록해야만 해결되는 문제이다.

3. 2 검색엔진

검색엔진이라는 용어는 웹으로 제공되는 문서들을 대상으로 검색할 수 있는 서비스를 의미한다. 그러므로 검색엔진 서비스의 검색결과는 사이트의 목록이 아니라 특정한 문서들의 목록이 되어야 한다. 그러므로 검색기능을 제공하고 있는 목록서비스 중 상당수가 검색엔진서비스로는 취급되지 않는다. 예를 들어 가장 유명한 목록서비스인 “야후”의 경우 검색기능을 이용해서 검색을 하면 그 결과로 웹사이트의 목록을 볼 수 있다. 그러나 이것은 검색결과가 문서의 목록이어야 한다는 검색엔진 서비스의 정의에 어긋나므로 “야후”와 같은 목록서비스는 검색엔진으로 보지 않는다. 검색엔진서비스는 모든 웹사이트의 정보를 대상으로 검색하고, 그외의 서비스들은 자신의 사이트 내부를 대상으로 검색한다고 생각하면 된다. 만약 어떤 검색엔진서비스와 같은 소프트웨어를 사용하는 사이트일지라도 검색결과가 웹문서가 아니라면 검색엔진이라고 보지 않는다. 즉 사용자가 원하는 자료를 찾을 때 사용되어지는 검색사이트 자체에 검색알고리즘을 가지고 있는 프로그램이 존재하는데, 이 프로그램을 지칭하여 검색엔진이라고 부른다.

따라서 본 연구에서는 검색엔진 뿐만 아니라 목록서비스도 많이 이용되고 있기 때문에 검색도구라는 용어로 두가지 서비스를 모두 포함시키되 각각 분리해서 평가하고 있다. 하지만 요즘 통용되는 말로

는 검색을 지원해 주는 웹사이트 자체를 검색엔진이라고 부르기도 한다.

3. 2. 1 검색엔진의 유형

하나의 데이터베이스에 모든 URL을 저장하고 특정 단어(키워드 또는 주제어 또는 검색어)를 입력함으로써 원하는 정보는 찾는 방법이다. 간단한 키워드를 통하여 원하는 정보를 신속하게 찾을 수 있는 장점이 있는 반면에 단점으로 색인이 정확하지 않은 검색엔진의 경우 원하는 정보를 찾을 수 없는 경우도 있다. 즉 해당 검색엔진이 얼마나 많은 자료를 데이터베이스에 정확하게 유지 및 관리하고 있는지가 중요한 관건이다.

국내의 대표적인 단어별 검색엔진은 심마니, 까치네, 정보탐정, 유니파인더, 애니서치, 웹글라이더이며 그밖에 와카노와 마당발 등이 있다. 본 장에서는 대표적으로 가장 많이 사용되는 심마니, 까치네, 정보탐정, 유니파인더, 애니서치, 웹글라이더 등에 대해 분석하고자 한다.

3. 2. 1. 1 단어별 검색엔진의 유형

1) 심마니 (<http://simmany.hnc.net>)

→ 뛰어난 자연어 처리

심마니는 이제까지 개발된 국내 검색엔진 중 가장 강력한 기능과 방대한 데이터베이스를 자랑하고 있다. 한글과 컴퓨터사의 자연어 처리팀에서 개발한 것으로 다양한 검색옵션을 제공하며 사용이 편리해 이용자가 많다. 영어권에서 만들어진 검색시스템과는 달리 한글검색을 효과적으로

가능하게 해 준다는 평을 받고 있다. 즉 심마니의 가장 큰 장점이라고 할 수 있는 자연어 처리기능은 한글의 특징적인 면을 고려해서 검색을 수행하는 뛰어난 검색로봇을 가지고 있기 때문에 한글검색에서 효과적인 것으로 보인다(마이크로소프트 웨어 96년 4월호, 월간 뉴미디어 96년 6월호, 컴퓨터 월드 96년 4월호, MYCOM 96년 5월호, HOME PC 96년 4월호, PC LINE 96년 4월호, PC PLUS 96년 6월호, PC WORLD 96년 4월호). 따라서 복잡한 검색연산자를 몰라도 흔히 사용하는 언어를 직접 입력해 검색결과를 얻을 수가 있다. 즉 사용자가 연산자 사용의 불편함을 느끼지 않도록 자연어 불리언 연산자와 한글 외래어를 영문으로 변환하여 검색하는 인공지능 검색까지 지원하고 있다.

화면하단에는 다양한 심마니 서비스가 제시되어 있는데, 특히 고급 검색을 지정할 수 있는 “나만의 심마니”와 적절한 비용을 지불하면 개인 혹은 기업의 홈 페이지에서 검색엔진으로 동작할 수 있도록 한 것이 주목된다. 특히 “나만의 심마니”는 검색을 하다보면 어느 정도 정보의 출처를 알고 있는 경우 특정한 곳만을 검색하여 결과물의 양을 줄이고 싶은 경우가 있을 때 이용하면 편리하다. “나만의 심마니”는 직접 원하는 검색 호스트를 한정할 수 있는 기능이다. 서버와 도메인이라는 입력조건을 기입하면 된다. 예를 들면 국내 교육기관에서만 자료를 찾고 싶다면 도메인에 ‘ac.kr’을 지정해 주면 된다. 물론 검색하는 자료는 이미 보유하고 있는

데이터베이스에서 검색조건에 맞는 내용을 골라내는 것이다.

한편 심마니는 로봇배제를 위한 제안 표준안을 지키고 있다. 로봇배제를 위한 제안 표준안이란 로봇의 방문을 원하지 않을 때, 해당 호스트에 /robots.txt 라는 파일을 만들어 로봇접속을 거부하는 경로를 명시하는 것이다. 그러면 해당 로봇은 이 파일을 먼저 참조해 검색대상 경로가 배제된 경로이면 접속하지 않고 다음으로 넘어가므로 인터넷 자원이 절약된다. 현재 인터넷에는 수많은 수의 로봇이 떠다니고 있기 때문에 이러한 로봇은 주요한 소스의 과부하 요인이 되기도 한다.

또한 심마니에서는 국내외의 몇가지 대표적인 검색도구를 모아 심마니를 통해서 다른 검색엔진으로 연결될 수 있도록 하고 있다. 만약 심마니에서 원하는 정보를 찾지 못했다면 이곳에서 다른 검색엔진으로 이동해 보는 것이 좋을 듯하다. 연결된 국내 정보검색시스템으로는 마당발, 까치네, 미스 다찾니, 인터넷 정보탐정, 애니서치, 코식크, SETSEARCH, iBOAT, SEARCHISLAND, UNIFINDER, WAKANO, WEBGUILDER, ZIP, ZOOM 등이 있고 국외정보검색시스템으로는 Excite, IBM AQUA, INKTOMI, Magellan, MetaSearch, OpenText, Snoopie, Tribal, WWW Worm, Web Crawler 등이 있다.

- 2) 까치네 (<http://kachi.tague.ac.kr>)
→ 도메인 검색이 가능

까치네는 대구대학교 학생들이 로봇을 이용해서 만든 검색엔진으로서 처음 서비스는 1996년 1월 10일부터 시작되었다. 그러나 까치로봇이 '로봇배제'를 위한 표준을 충분히 만족시키지 못하고 논리연산자를 사용할 수 없을 뿐만 아니라 출력되는 결과가 간단한 순이어서 까치네2로 새롭게 단장하게 되었다. 현재의 "까치네"는 대구대학교 정보통신 공학부 학생들의 컴퓨터 기술과 응용미술과 학생들의 그림그리기 기술로 만들어진 것으로 유즈넷 검색과 방명록 역할을 하는 쪽지와 도메인 검색을 지원하는 색다른 기능을 추가해서 만들어졌으며, 최근에는 분류서비스도 실시하고 있다. 사용방법이 평이하고 화면이 아기자기해서 복잡한 심마니와는 대조적이며 50만건이 넘는 많은 데이터베이스를 확보하고 있다.

까치네의 메뉴들은 이미지 맵을 통해 제공되고 있는데 주요한 메뉴들에 대해 간단히 설명해 보면 다음과 같다. 먼저 까치네의 독특한 서비스인 "도메인별 검색"을 살펴 보면, 기본적으로는 검색서비스와 같은 방법을 사용하는 것인데 다만 검색대상을 특정 도메인으로 축소시켜 더욱 정확한 검색결과를 얻을 수 있도록 해주고 있다. 이 서비스를 자신의 각 도메인의 중요 호스트에 넣으려면 미리 도메인 값에는 자신의 도메인을 HIDDEN으로 넣은 다음 검색어만 입력 받으면 된다. 그러나 "나만의 심마니"와 비교해 볼 때 많은 검색 옵션을 제공하지 않으므로 검색어가 복잡하다든가 옵션이 많을 경우 대부분

관련된 홈페이지를 찾을 수 없다는 메시지를 받게 된다. 다음으로 독특한 서비스인 "URL찾기"를 살펴 보면, 입력된 URL을 가지고 인터넷상에 이 URL을 링크하고 있는 사이트들을 알려주는 서비스라는 것을 알 수 있다. 즉 자신의 URL을 입력해 봄으로서 자신의 URL이 국내 인터넷계에 어느 정도 알려져 있는지 가능해 볼 수 있고 URL 진단을 통해 해당 URL이 링크하고 있는 페이지에 오류가 없는지, 링크가 잘못 연결되어 404 NOT FOUND는 없는지를 확인해 준다. 또한 "싱싱나라"에서는 URL과 EMAIL을 입력해 두면 해당 URL의 내용이 바뀔 때마다 지정된 주소로 까치네 로봇이 해당 URL의 날짜를 비교해서 메일을 보내준다.

검색대상은 국내 웹으로 한정되어 있고 화면하단에서 보듯이 본문과 제목 등의 옵션으로 검색 할 수 있다. 본문 전체 검색은 웹문서의 본문 전부를 제목 검색은 웹문서의 제목만을 각각 검색한다.

3) 정보탐정 (<http://aistar.kotel.co.kr>)

→ 어구탐색이 가능한 정보탐정

정보탐정이란 한국통신 멀티미디어 연구소에서 자체 개발한 정보검색엔진을 사용하여 만든 인터넷 서비스 시스템으로서 웹페이지 검색, 유즈넷 뉴스 검색, 웹페이지 분류, 광고 등의 서비스를 제공하고 있다. 현재 웹페이지 검색은 국내 문서와 국내에서 연결된 관련 해외 문서(매주 점검하고 색인한다)로 한정하고 있으나 앞으

로 전세계 문서 검색으로 대상을 확장해 나갈 방침이다.

현재 사용되고 있는 것은 정보탐정Ⅱ로써 이보다 앞서 1996년 3월 정보탐정Ⅰ이 첫 서비스를 개시 이래 각종 서비스를 추가해서 개량된 버전이다.

정보탐정Ⅱ의 주요 특징들 가운데 시스템 구성상의 가장 큰 장점은 문서수집 로봇과 색인기, 검색기, 자연언어 처리기, 인티페이스들을 각각 완전히 독립시켜 분리해 둘 수 있다는 점이다. 특히 문서수집 로봇과 색인/검색기는 많은 양의 계산을 소요하기 때문에 서로 다른 호스트에 따로 분산시킬 필요가 있다. 또한 HTTP 서버와 검색서버를 분리하여 부하를 분산시킬 수 있는 구조로 설계되어 있다.

또한 정보탐정은 시그너처 기법을 이용한 고속 색인 및 검색을 수행하는데 색인 검색 프로그램은 100만개 이상의 문서까지도 수용할 수 있다. 시그너처란 입력 스트링을 특정한 해싱함수(hashing)를 이용해서 만들어 내는 고정길이의 비트열이며 시그너처를 이용한 색인이란 어떤 문서를 하나의 시그너처 비트열로 만들어 내는 작업을 반복적으로 수행하는 것을 말한다. 시그너처의 폭은 고정적인 크기이므로 입력문서의 개수만 알면 필요한 시그너처 색인 파일의 크기는 산술적으로 계산할 수 있다. 시그너처 색인을 만드는 것은 다른 요소와는 무관하게 문서의 개수에만 선형적으로 증가하기 때문에 대량 문서를 다루는데 적합하다. 그리고 시그너처를 만드는 과정이 단순한 산술적인 해

싱함수를 사용하는 것이므로 견고한 색인기를 만들 수 있다. 그러나 false drop checking 작업으로 검색속도가 떨어지는 문제점을 안고 있는데 정보탐정은 상위 수십개에 대해서만 점검하기 때문에 속도가 크게 떨어지지 않는다. 그리고 문서수집 로봇은 중복됨이 없이 모든 문서를 다운로드하되, 다음에 똑같은 문서를 두 번 다운로드 하지 않도록 하여 네트워크의 부하를 최소화하도록 하고 있다.

정보탐정에서는 시그너처 색인구조를 쓰고 있기 때문에 색인시에는 자연언어 처리기를 사용하지 않고 검색시에만 자연언어 처리기를 사용하는데 자연언어 처리기의 역할은 1)명사추출, 복합명사 분할 2)시소러스, 유사어 대치 3)블리언 연산자 사용인데 자세한 검색방법은 뒤의 정보검색법에서 언급된다.

뿐만 아니라 국내 문서에서 나타나는 한글, 영문, 한자 모두를 검색하고 추가 작업이 없이도 아시아권 문서를 포함한 전세계 문서를 색인하고 검색할 수 있어 전세계 시장으로 확대할 수 있다. 웹페이지 뿐만 아니라 DBMS의 데이터, 정형, 비정형의 각종 TEXT 데이터 등 모든 형식의 데이터를 문제없이 처리할 수 있는 검색엔진이다. 마지막으로 정보탐정을 이용하여 만들어진 주요 상품은 정보통신문헌정보시스템과 한국통신연구업무 편람 검색시스템과 한영 성경 검색시스템이다.

4) 웹글라이더

(<http://www.infoglider.com>)

→ 다양한 연산자 지원

웹글라이더는 삼성 데이터시스템 정보 기술 연구소의 TextGlider를 기반으로 이루어진 검색엔진이다. 초보자는 물론이고 고급 검색가에게까지도 효율적으로 사용할 수 있도록 다양한 검색옵션으로 위치연산자 기능을 지원하고 있을 뿐만 아니라 구문, 절단, 필드, 도메인, 깊이검색 등을 갖추고 있으며, 객체저장시스템기술, 로봇 에이전트 기술등 최신기술을 이용한 것이 특징이다(동아일보, 전자신문, 중앙일보, 한국경제신문 각 97년 2월 27일, 경향신문 97년 2월 26일, 매일경제신문 2월 26일). 또한 형태분석기를 사용하여 정확한 한글 한자 및 영어처리가 가능하도록 하고 있다. 따라서 자동 한글 키워드 추출, 한자어에 대한 별도 색인, 영어의 STEMMING처리가 이뤄지도록 해서 검색결과가 신뢰할 만하다. 그리고 검색결과를 출력해 줄 때는 문서 가운데 입력된 검색어와 가장 관련이 있는 문장을 동적으로 추출하여 하이라이트하여 보여준다. 그러나 검색어에 따라서 문서에 대한 요약 정보가 아닌 키워드와 같은 짧은 구들이 나타나는 경우도 있어 문서의 내용을 파악하기가 어려울 경우도 있다.

웹글라이더는 해외유명검색엔진인 "MetaGlider"를 통해 통합검색엔진의 서비스도 실시하고 있다. 이 두 서비스는 검색엔진의 구조뿐만 아니라 검색대상 및 검색옵션이 서로 다르므로 검색어 입력 방법 또한 서로 다르므로 사용자가 검색시 웹글라이더에서 사용한 검색어 입력방

법으로 메타글라이더를 적용하면 원하는 결과를 얻지 못하게 된다. 그러므로 각 서비스에 맞는 방식을 선택하여야 한다. 메타글라이더가 검색을 의뢰하는 검색엔진은 Alta Vista, Infoseek, Excite, Open Text, WebCrawler, Yahoo 등이며 모든 검색엔진들로부터 받은 검색결과를 재정리하여 순위와 가중치를 부여하여 제목, URL, 요약정보들을 화면에 보여준다.

5) 유니파인더

(<http://www.unifinder.com>)

→ 유니텔의 베타검색 서비스

유니파인더는 삼성전자 멀티미디어 연구소(유니텔)에서 개발한 한글, 영문 정보 검색엔진이다. 30만건이 넘는 데이터량으로 국내웹과 유즈넷 메타서치를 이용한 국외사이트를 검색해서 해외문서를 검색해 주며, 신문검색도 가능하다.

문서들을 자동으로 색인하며 이미 만들어진 색인 데이터베이스에 새로 수집되는 문서를 즉시 추가하는 것이 가능해 최신 정보를 저장하고 검색하기에 좋다. 특히 유니파인더는 삼성 종합 기술원이 가지고 있는 뉴스서버를 이용해서 유즈넷 문서를 색인하므로 유즈넷에 올라와 있는 토론자료나 새로운 기사를 읽고자 할 때 좋다. 또한 밤사이에 신문기사를 수집하므로 아침신문을 검색할 수도 있다.

한글과 영문등 각 언어의 특성에 맞는 방법으로 색인하기에 한글과 영문 검색어에 맞는 정보를 찾을 수 있다.

6) 애니서치

(<http://www.anysearch.com>)

→ 가상공간의 비즈니스 서비스

삼성전자가 개발해서 삼성물산에 공급한 것으로 W3COM의 표준기술을 적용한 로봇을 이용해서 웹사이트의 정보를 자동적으로 수집하여 코다 이지베이스 등의 자연어 처리능력을 갖추고 있다(매일경제 97년 3월1일, 전자신문 97년 3월 3일). 삼성물산은 이 서비스의 활성화를 위해 광고대행사인 (주)거손과 공동으로 마케팅 계약을 체결, 광고유치와 홍보에 나서고 있다. 특히 이용자들이 다른 사이트를 보는 동안에도 해당광고에 노출될 수 있는 동영상 효과를 구현하는 프레임 광고 기능을 이용해서 온라인 광고효과를 높이고 있다. 한국판 야후를 꿈꾸는 이 서비스가 관심의 초점이 되는 것은 검색서비스보다는 바로 이러한 전자상거래 등 가상공간 비즈니스의 첫걸음을 내딛고 있기 때문이다(중앙일보 97년 3월 10일). 그러나 문헌정보학적인 측면에서 무엇보다도 관심을 끄는 것은 문헌정보학 전문가들에 의해 세밀한 주제 분류서비스가 구현되었다는 점이다(정보통신신문 97년 3월 3일). 대다수의 검색엔진들이 비전문가에 의해 주제가 분류되어 각 검색엔진들마다 서로 다르게 주제분류가 이루어져 있기 때문에 주제분류 서비스로 접근해서 정보 검색을 할 경우에 찾고자 하는 주제가 담긴 카테고리를 찾는 것만으로도 많은 시간을 낭비할 경우가 종종 있다. 그러나 문헌정보학 전문가들의 색인분류학에 기초

한 체계적인 분류들로 모든 검색엔진들이 이루어진다면 훨씬 더 빠르고 쉽게 원하는 정보에 접근할 수 있을 것으로 생각된다. 또한 접속 시간의 절약은 비용 절약과도 맞물려 있으므로 시간 대 비용에 있어서도 더욱 효율적인 검색이 이루어질 수 있다고 본다.

애니서치의 가장 큰 특징은 신문통합검색기능을 가지고 있다는 것이다. 따라서 국내 일간지들의 모든 기사를 하나의 키워드만으로 찾아볼 수 있는 잇점을 가지고 있다. 웹문서와 유즈넷 그리고 신문기사검색이 가능하다.

3. 2. 1. 2 메타검색엔진

메타(Meta)검색엔진이란 많은 검색엔진들을 모아 검색하기에 편리하게 만든 툴로써, 웹 자원뿐만 아니라 Software, People, Publication, News/FAQs, Documentation 등의 환경까지도 제공한다. 즉 웹문서만을(HTML문서) 검색대상으로 하는 것이 아니라 프로그램이나 뉴스그룹 또는 FAQ같은 문서까지도 찾아주는 아주 넓은 검색영역을 갖는 것이 특징이다. 따라서 각각의 검색엔진을 옮기면서 질의할 필요없이 하나의 조작환경(interface)으로 각각의 검색엔진을 검색할 수 있는 장점을 갖는다.

1) 미스 다찾니

(<http://www.mochanni.com>)

→ 지능형 메타검색엔진

미스 다찾니는 삼성전자 멀티미디어 연

구소의 이승현씨에 의해서 개발된 것으로 자체적인 로봇이나 데이터베이스를 가지고 있지 않고 다른 검색엔진들의 검색결과를 한 곳에서 모아 보여주는 이름 그대로의 한국적인 지능형 메타검색엔진이다. 국내의 대표적인 메타검색엔진들로는 심마니, 까치네, 정보탐정, 마당밭, 코식크, 메타글라이더, 유니파인더 등인데 이러한 메타검색엔진 가운데 “미스 다찾니”의 가장 큰 장점은 한 번의 검색으로 국내외의 여러 다양한 검색엔진을 이용하는 효과를 얻는 통합검색엔진의 성격을 띠는 것이다. 그리고 각 검색엔진에서 얻은 결과물은 미스 다찾니의 자체 순위 결정 알고리즘에 따라 관련성이 높다고 생각되는 정보를 먼저 출력하고 있는데, 중복된 정보는 통합하여 보여 준다. 구조는 클라이언트 지향 분산 구조에 따라 에이전트를 실행하는 서버 및 네트워크의 부하를 줄이도록 하고 있으며, 각 사용자에게 대한 응답속도를 향상시키고 있다.

검색영역은 웹과 신문으로 웹에서는 영어검색엔진과 한글검색엔진을 모두 이용할 수 있다. 대부분의 신문 사이트들은 메인 페이지에서 검색기능을 제공하지 않고 있다. 따라서 기사검색을 하려면 버튼을 여러번 클릭해야 하는 번거로움이 있다. 그러나 미스 다찾니를 이용하면 간단하다. 특히 기사검색은 다양한 연산자가 필요없기 때문에 더욱 유용하게 각 신문사들의 데이터베이스를 한번에 검색할 수도 있다.

미스 다찾니는 단순한 화면 디자인 만큼 사용법이 간단하고 메뉴가 단순해서

쉽게 접근할 수 있으며 한 번의 검색어 입력으로 국내외의 주요 검색엔진과 신문 사이트로 연결될 수 있어 검색범위가 방대해서 국외의 메타크롤러와 같은 메타검색엔진의 특징을 갖추고 있지만 각 검색엔진의 고유한 검색방법을 사용할 수가 없기 때문에 효율성이 떨어질 수 있다. 다시 말해서 정보검색의 범위는 방대하고 결과물도 많지만 정밀한 검색은 어려울 것으로 보인다(정확율의 저하). 그러나 이러한 특징들이 곧 메타서치엔진의 장단점이 되기도 한다.

이 검색엔진은 하나의 데이터베이스만을 검색하는 것이 아니고 각각의 검색엔진들에 접속해서 이용자가 입력한 검색어로 여러 검색엔진들을 검색하기 때문에 네트워크 및 해당 검색엔진들의 부하에 따라 검색속도가 달라진다. 따라서 너무 속도가 느린 검색엔진 때문에 다른 검색엔진의 결과를 보지 못하는 처리시간상의 문제가 있을 수 있다. 그래서 제한된 시간 안에 처리된 검색 결과만을 출력되도록 하는 옵션을 제공한다. 30초, 1분, 2분, 3분, 4분 등 네가지 옵션이 있다. 그러나 만약, 검색시간을 30초로 설정했는데, 몇군데의 결과밖에 얻지 못했다면 처리시간을 늘려주고 검색을 다시 한 번 해 보는 것이 좋다. 이미 가져온 결과는 캐쉬에서 찾아내고 못가져 온 검색엔진으로만 검색을 의뢰하는데 캐쉬의 유효시간은 2시간이다.

검색결과 중 ‘네트워크로 접속이 안됩니다.’라는 에러 메시지가 나오면 프록시를 사용해서 검색하는 것이 좋은데 여기

서 프록시를 사용하면 다른 경로로 해당 검색엔진에 접속해서 검색을 하게 된다. 이외에도 출력 옵션에는 자세히, 간단히, 보통 등이 있는데 '보통'과 '자세히'는 거의 비슷하고 '간단히'는 해당 페이지의 제목과 찾아낸 검색엔진들만이 출력된다. 결과 옵션에는 통합과 비교가 있다. '통합'은 여러개의 검색엔진이 같은 페이지를 찾았을 경우, 하나의 결과만 나타나지만 '비교'는 각각의 검색엔진에서 찾아낸 검색 결과를 모두 출력한다. 또한 순위를 결정하는 옵션이 있는데, 순위결정이란 입력된 검색어를 많이 포함한 페이지를 우선으로 출력하는 것이다. 순위결정 옵션에는 지능적, 간단히, 없음이 있다. '지능적'이란 순위를 정해서, '없음'은 순위없이 출력하는 것을 말한다.

미스 다찾니는 별도의 연산자 지원은 하지 않고 기본적으로 퍼지검색을 적용하고 있는데, 퍼지검색은 불리언 검색과는 달리 키워드의 존재 유무를 0에서 1사이의 값으로 표현하여 보다 부드럽고 기능적으로 순위결정을 해주는 가중치 기법을 사용하고 있으며, 방법은 찾고자 하는 정보에 필요한 단어들을 띄어 쓰면 된다. 특정한 어떤 기호가 필요하지 않다. 꾸준한 개선이 이루어진다면 한국적인 메타서치 엔진으로서 국외의 "MetaCrawler"에 견줄 수 있는 그런 검색엔진이 될 것이다.

미스 다찾니에서 사용할 수 있는 검색 도구로는 다음과 같은 것들이 있다.

먼저 국내 검색엔진으로는 미스 다찾니, 마당발, 까치네, 코식크, iBOAT, 유니파인

더, WAKANO, 웹글라이더, 정보탐정, 심마니, ZOOM 등이며 국외 검색엔진으로는 Alta Vista, Infoseek, Lycos, Yahoo, DejaNews, Excite, Insane Search, Mama, OnRamp, Savvy Search, MetaCrawler, HotBot 등이다. 그외에도 뉴스로는 서울신문, 한국일보, 한겨레신문, 매일신문, 문화일보, 중앙일보, 조선일보, 동아일보, 경향신문, 한국경제신문, 매일경제신문, 전자신문, CNN, LA Times, USA Today, 워싱턴포스트지, Yahoo News, News.Com, ThecWeb, WiReD, Excite Live, ZD Net News 등이고 그밖의 사이트로는 한국관광사이트, 인터파크, 롯데백화점, 해외잡지, 국내잡지, (주)정보시대, 기업명단, 씨티넷, 비조넷, 넷퀘스트21, 왕십리분원, 초보자를 위한 HTML 가이드, 인터넷 정보검색세미나, Unix Guru Universe, Unix Reference Desk, Computer Science Resources 등이 있다.

3. 2. 1. 3 통합검색엔진

한 번의 검색으로 여러 검색 엔진을 동시에 검색, 대부분 멀티쓰레드 (MultiThread) 기법을 사용하므로 한 번의 검색시간으로 여러 곳을 검색할 수 있다. 국내에서는 미스 다찾니가 메타검색엔진인 동시에 대표적인 통합검색엔진이다. <멀티쓰레드 질의 페이지 : URL → <http://www.sun.fi/mtq/mtquery.html>>

3. 2. 2 검색엔진의 정보검색법

목록서비스를 이용한 정보검색이 특정한 주제를 가지는 사이트를 찾는 데 유용하다면 검색엔진은 특정한 주제를 가지고 있는 '문서'를 찾기에 유용한 것이다. 검색엔진은 메뉴를 선택하듯 주제분류를 선택하는 목록서비스보다 더욱 강력한 검색능력을 제공하는 정보검색도구인데 이는 아래와 같은 정보검색 옵션들을 사용할 수 있기 때문이다. 먼저 검색법 작성에 관한 주요 옵션들을 설명한 다음 각 검색엔진들의 정보검색방법들에 대해 설명하고자 한다.

1) 부울연산자 (Boolean Connectors)

검색도구에 따라서는 검색식에 직접 연산자를 쓰지 않고 옵션으로 선택하게 되어 있는 경우(Yahoo)와 연산자를 쓰거나 선택하는 일이 없이 기본으로 and나 or연산(Lycos, 코식크)이 되는 경우가 많이 있다. 그러나 일반적으로 표준 검색방식에서는 다음과 같은 연산자들을 검색식내에, 검색어 바로 앞에 공백없이 직접 입력한다. 그리고 연산자는 대개 소문자로 입력한다.

(1) and

연산자 좌우의 검색어가 모두 나타나는 자료를 찾아준다. 주로 and나 ' + ' 로 표시하지만 연산자 없이 그냥 공백만으로 and 연산을 하거나 그외 '&' 기호를 쓰는 경우도 있다. 심마니에서는 '그리고/와/과/및' 등의 한글 연산자를 사용할 수 있다.

(2) or

연산자 좌우의 검색어 중 어느 하나만이라도 들어있는 자료를 찾아준다. 주로 공백이나 or를 사용하지만 ' + ' 또는 ' , '를 사용하기도 한다. 검색도구에 따라서는 적합성 점수(가중치)를 부여할 때 or 연산자의 좌우의 단어가 모두 들어있는 자료에 높은 점수를 주기도 한다.

(3) not 또는 and not

연산자 앞쪽의 검색어를 포함하되, 뒤쪽의 검색어는 들어가지 않는 자료만을 찾는다. 주로 ' - '나 not를 많이 사용하지만 '! ' '&!'를 사용하기도 한다.

2) 구절검색 (Phrase Searching)

두개 이상의 단어가 순서대로 연속해서 나오는 것을 찾아준다. 보통 구절을 " "로 감싸주는데 검색엔진은 이 구절을 하나의 긴 문자열로 취급하는 것이다. 주로 두 단어 이상으로 되어 있는 전문용어 들을 찾을 때 사용하지만, 검색 문자열에 특수기호(. , /)나 불용어까지 포함시키려는 경우에도 사용한다. 구절검색을 할 경우 절단검색 기호(*)는 사용할 수 없고 불용어가 아닌 단어를 두개 이상 포함해야 한다.

3) 검색식의 문법 (Grouping by Parentheses)

인터넷의 검색도구들에서도 대체로 여러개의 검색어와 연산자들을 결합하는 복잡한 검색식을 사용할 수 있다. 그러나 대개는 상용 데이터베이스와는 달리 and 연산자가 or 연산자보다 우선되는 것도 아니

고 그저 검색식의 왼쪽으로부터 오른쪽으로 차례대로 연산하는데, WebCrawler, Alta Vista, 유니파인더, 심마니, 애니서치 처럼 검색연산순서를 바꿀수 있는 것이 있다.

4) 인접연산 (Proximity Searching)

인접연산자는 앞뒤의 검색어가 동시에 나와야 한다는 점에서 and 연산과 비슷하지만 두 검색어가 바로 옆에 또는 일정 거리내에 인접해 있어야 한다. 한편, 앞 뒤 검색어가 얼마나 가까이 붙어 있어야 하는가, 두 검색어 사이에 몇 개까지의 단어가 끼어드는 것을 허용하는가 하는 인접 정도는 검색도구에 따라 조금씩 다르다. 100단어 이내(Infoseek), 5글자이내, 또는 한 문단내에 인접해 있는 조건으로 해석되기도 하는데 국내 검색엔진으로는 유일하게 웹글라이더가 2개의 문자까지를 허용하고 있다.

인접연산자는 보통 하나의 검색식 내에서 하나만 사용할 수 있다. 여러개의 인접연산자로 구성된 검색식은 거의 허용되지 않고 그렇게 써야 할 경우도 거의 존재하지 않는다. 또한 구절 검색과 인접연산자를 함께 사용할 수도 없다.

- (1) 검색어간의 순서까지를 지정하는 ADJacent, Preceding, ~

연산자 좌우의 검색어가 지정된 순서로 인접해 들어 있는 자료를 찾아준다. Dialog의 Within연산자와 같다.

- (2) 검색어간의 인접관계만 지정하는 Near, ^

이는 좌우 검색어가 앞뒤 순서에 상관 없이 인접해 있는 경우를 말한다. Dialog의 Near연산자와 같다.

Near 사용 : WebCrawler, Alta

Vista ^ 사용 : 웹글라이더, 데자뉴스

5) 절단검색

검색식에 입력된 문자열로 시작하는 단어는 모두 검색해 준다든가 하는 기능을 절단검색이라고 한다. 절단검색에는 검색어 뒷부분의 변화를 허용하는 우절단을 비롯하여 앞부분의 변형을 허용하는 후방일치검색 등이 있다.

검색용어의 끝부분을 임의의 문자로 지정해 줌으로써 영어의 경우에는 주로 단수형과 복수형을 한꺼번에 검색하거나, 품사의 차이로 어미변화를 일으키는 용어를 검색할 때 사용하고, 우리말과 일본어의 경우 앞이 같은 말로 시작되는 유사어나 복합어(장애, 장애인, 장애인복지, 장애인, 장애인시설, 장애인연금)를 검색하고자 할 때 이용한다.

절단검색어를 사용할 때 어간이 짧은 경우에는 관계없는 용어들이 같이 검색되므로 어간을 너무 짧게 지정하지 않도록 주의해야 한다. 영어의 경우 communication, communicate 등 의사소통에 관한 자료를 찾기 위해 검색식을 com*이라고 쓰면 community, computer, compass, coma, commercial ... 등 전혀 원하지 않는 용어들까지 검색하게 된다.

- (1) 전방일치검색 단어* (또는 %단어)

절단검색 옵션을 선택하거나 검색어 뒤에 절단기호를 붙인다. 절단기호로는 대문분 *를 사용하지만 임의의 글자 수만큼 ?를 사용하는 경우도 있다. 한편 검색식에서는 절단기호를 사용하지 않고 우절단검색, Word-stem based search, Substrings, Start with, 전방일치검색 등으로 표시한 옵션을 선택하게 되어 있는 검색시스템들도 많다. 단순히 절단검색이라고 하면 이 우절단검색을 의미하는 것이라고 할 수 있다.

(2) 후방일치검색 *단어 (또는 단어%)

대부분 *를 사용하지만 임의의 글자 수만큼 ?를 사용하는 절단방식도 있다. 인터넷상의 거의 모든 검색도구들이 후방일치절단검색을 지원하지 않는다. 그러나 영국의 웹검색기인 WWLib 'Extended Search Interface'에서는 이 기능을 지원한다. 국내 검색엔진인 정보탐정의 경우 검색식을 '장애%' 라고 쓰면 지체장애, 청각장애, 시각장애 등 장애로 끝나는 용어들을 찾아준다. 검색용어의 시작부분을 임의 문자열로 지정하는 것으로서, 영어단어일 경우 주로 화합물의 검색을 위해 사용하는데 위의 예와 같이 이 검색식은 한글 복합어를 검색할 때 특히 유용하다.

(3) 사이에 끼는 가변 글자 단어?단어

Dialog와 같은 곳에서 주로 사용하는 것인데 인터넷상의 정보검색엔진에서는 이런 검색방식을 지원하는 경우가 거의 없다.

6) 필드제한검색

웹문서의 제목, 본문, 주소 (title, body, address)등이나 뉴스그룹 기사의 저자, 주제, 뉴스그룹이름(author, subject, newsgroup name) 등 문서의 특정 부분으로 제한하여 검색하면 보다 정확하고 빠르게 검색할 수 있다.

(1) 검색식에 직접 필드제한자를 붙이는 경우

가장 유연한 방식으로써 고급 검색가들에게는 검색식내에 직접 여러 연산기법들을 사용할 수 있게 하는 방법이다. 국외의 Alta Vista의 경우 필드제한검색 특히 title, url, link 필드에서의 검색은 정확한 자료를 찾아내는데 큰 도움이 된다. 그리고 이것은 예전에 책갈피에 만들어 둔 좋은 자료가 url의 무단변경으로 인하여 더 이상 찾아내기 어려울 때 아주 유용한 검색 방식이다. 예를들면 url:spcr+title:"Special Policy Research Centre"의 검색은 현재 정확한 주소는 모르지만 url과 타이틀 중에 포함된 단어로써 찾아내는 것이다. 자신의 홈페이지나 웹문서를 다른 곳들에서 얼마나 많이 링크하고 있는지를 알고 싶을 때도 사용할 수 있는데 데자뉴스에서 가능하다.

(2) 검색식을 입력한 다음 검색대상 필드를 옵션으로 선택하는 경우

Galaxy나 OpenText 파워 서치를 비롯한 대부분의 국내 검색엔진들 가운데 가치네, 웹글라이더 등이 이에 해당한다. 단도메인 검색을 지원하는 경우는 (1)의 방식을 지원한다.

(3) 필드별 검색식 입력상자에 검색식

을 작성하는 경우

데자뉴스 쿼리필터(Daja News Query Filter), REHABDATA, 유럽사회과학 정보협회의 통합검색 CESSDA 등이 해당된다.

7) 불용어

불용어란 검색엔진이 검색에서 무시해 버리는 문자열을 의미한다. 영어의 전치사나 관사 등이 의미없는 단어로 취급되어 검색어로 사용되지 않는 것인데 국내 검색엔진의 경우에는 특수기호(. , +) 등이 해당된다. 그러나 이러한 특수기호가 필요할 경우도 있는데 구절검색기능을 이용하면 된다. 국외 검색엔진의 경우에는 구절 검색으로 특수기호를 지원하지만 국내 검색엔진의 경우에는 아직까지는 미흡하다.

8) 집단검색어, 유의어, 시소러스, 확장 검색

(1) 집단검색어

예를 들어 검색을 '대도시'라고 썼을 때 전국의 대도시 지명들을 OR조건으로 모두 나열한 것과 같은 효과를 내는 검색 방식이다. 이것은 집단검색어 파일을 따로 갖고 있는 데이터에서만 사용할 수 있다.

(2) 유의어

유의어 사전을 지원하는 검색엔진이 있는데 대표적인 것이 국내의 심마니이다. 심마니에서는 예를들어 장애인 #이라고 검색식을 입력하면 장애인 뿐만 아니라 장애자, 장애인, 장애인 등도 찾아준다. 이와 같이 사용자가 검색식에 입력한 문자

열과 같지 않더라도 의미상 같거나 관련이 있는 용어들을 모아 검색이 가능하도록 구축한 데이터베이스를 시소러스 또는 유의어사전이라고 한다.

(3) 확장검색

영국의 웹 검색 기인 WWWLib Extended Search Interface는 좀 더 특별한 확장검색을 지원한다. 즉 명사의 복수형과 동사의 분사형이나 진행형 그리고 관련되는 단어도 검색해 준다. 국내 검색엔진의 경우 확장검색이 가능한 것은 "검색엔진"이라고 입력하면 검색도구, 검색시스템 등 관련있는 용어들을 찾아 주는데 검색시간이 길어지는 문제점이 있다.

3. 2. 2. 1 각 검색엔진의 개별분석

위에서 설명한 정보검색방법들로 국내 각 검색엔진들의 정보검색방법을 살펴보면 아래와 같다.

1) 심마니

검색어에 대한 검색결과를 화면에 표시하는 쪽당표시의 기본값은 10개 인데 이는 한 페이지에 검색한 결과 중 열 개의 웹문서를 표시한다는 의미이다. 정할 수 있는 결과의 수는 5개, 10개, 15개, 20개, 25개, 30개 등이고, 검색한 결과를 출력할 때에는 간단히와 표준의 두 방식이 있는데 간단히는 검색한 결과 중 파일 이름만 출력하며, 표준은 검색한 결과 중 정확도와 요약 일부를 출력한다.

심마니의 특징인 유의어에 있어서는 정확한 단어를 입력하지 못하더라도 비슷한 의미의 다른 말까지 골라서 비슷한 정도

까지 표현해 주는 기능이다. 이 기능은 초보자에게 특히 도움이 되는 기능이며 확장방법에는 한글확장과 영문확장과 발음확장의 세가지 있다. 발음확장의 경우 비슷한 발음에 대한 질의어도 모두 함께 찾아준다. 외래어 같은 경우 맞춤법을 제대로 알기 어려운 경우가 많은데 정확한 발음을 몰라도 비슷한 유추를 통해 검색이 가능하다. 단 음절수가 3음절 이상인 단어에서만 효능을 발휘한다. (예 : 팬티움, 펜티움, 팬티엄)

(1) 부울연산자

두개 이상의 질의어를 반드시 포함한 문서를 검색하기 위해서는 ‘&’, ‘그리고’, ‘와’, ‘과’, ‘AND’, ‘;’, ‘및’을 사용하고 두가지 이상의 질의어 중 한가지라도 포함하고 있는 문서를 검색하기 위해서는 ‘+’, ‘또는’, ‘OR’, ‘혹은’을 사용한다. 그리고 두 개 이상의 질의어 중 한가지를 포함하고 있는 문서를 검색하고자 할 때는 ‘~’을 사용하고 문서가 포함되지 않기 위해서는 ‘!&’ 와 ‘&!’ 를 사용한다.

(2) 유의어, 동의어를 검색하기 위해서는 ‘#’를 사용한다.

(3) 절단검색으로 *을 사용한다.

(4) 검색연산의 우선순위는 ()를 통해 결정한다.

(5) 구절 검색은 “ ” 사용한다.

2) 까치네

기본적인 연산자를 지원하고 있으며 본문 전체검색, 제목검색, 주소검색들을 별

도로 지원하고 있으므로 웹의 경우 URL을, 유즈넷의 경우 EMAIL을 검색대상으로 주소검색을 사용한다면 적합한 정보를 얻는데 도움이 된다. 검색어는 평범하면서 쉬운 것을 사용하는 것이 검색의 효율을 높여준다.

(1) 부울연산자

기본적인 부울연산자를 사용하고 있다. 즉 AND, OR, NOT의 역할을 하는 +, 스페이스, !을 사용해 검색식을 완성할 수 있다.

‘정보+검색’와 같은 검색식의 경우 ‘정보’와 ‘검색’ 두 단어가 모두 포함된 문서를 검색한다. 이 때, +연산자의 좌우에는 공백을 꼭 들 필요는 없다.

‘정보 검색’과 같이 중간에 공란을 두었을 경우에는 공란은 OR로 인식하여 검색을 수행한다. 즉 ‘정보’ 또는 ‘검색’ 또는 ‘정보’와 ‘검색’ 두 단어가 모두 포함된 문서를 검색한다. ‘검색! 정보검색’과 같은 검색식의 경우 ‘정보검색’이라는 문자열을 가진 문서는 검색하지 않고 ‘정보’라는 단어만 포함된 문서를 검색한다.

3) 정보탐정

다른 검색엔진과 마찬가지로 너무 자세하고 긴 검색어(예, 전기용품 안전관리법에 의한 형식 승인 절차)라든지 너무 일반적인 검색어(예, 한국)으로는 검색효과가 없으며 긴복합명사의 경우에는 띄어쓰기를 해 주는 것이 좋다.

(1) 부울연산자

두 개 이상의 검색어 중 한가지라도 포

합하고 있는 문서를 검색하기 위해서는 +, or, | 를 사용하고, 두 개 이상의 검색어를 반드시 포함하고 있는 문서를 검색하기 위해서는 *, and, & 를 사용한다. 그리고 not의 역할을 하는 - 가 있다.

(2) 구절탐색을 위해 “ ” 사용한다.

(3) 절단탐색

문서의 시작과 끝에 특정 단어가 위치한 문서를 검색 할 수 있다. 그러나 그 기능을 충분히 활용하기 위해서는 검색어를 신중히 사용해야 한다. “검색어”로 시작되는 문장이 있는 문서를 검색할 때는 ‘% 검색어’, “검색어”로 끝나는 문장이 있는 문서를 검색할 경우에는 ‘검색어%’, “검색어”만 단독으로 나타나는 문서를 검색할 경우에는 ‘%검색어%’를 사용한다.

(4) 대소문자구별

대문자가 하나라도 들어 있으면 정확히 일치하는 것만 찾고 소문자로 이루어져 있으면 대소문자를 구분하지 않고 찾는다. Times만 찾고 싶을 때는 ‘Times’를 입력하고, TIMES만 찾고 싶을 때는 ‘TIMES’를 입력하며, Times, TIMES, times, tiMes들로 입력하면 ‘times’를 검색해 낸다.

(5) 인접어 검색

여러 단어가 일정 거리 이내에 존재하는 문서를 검색해 주는 기능으로 단어들의 순서를 인정하는 연산과 무시하는 연산 두가지가 있다.

(6) 검색어 확장

유사어나 개념어 수준의 확장검색을 지원한다.

(7) 필드검색

문서를 필드별로 구분하여 검색하는 것이 가능하고 필드간 연산도 가능하다.

(8) 제한검색

검색결과가 너무 많을 때 실마리를 추가하여 검색결과 만을 대상으로 다시 검색해 주는 기능을 제공한다.

(9) 대상 제한검색

정형화된 문서에 코드나 연도와 같이 구분된 정보가 들어 있는 경우 특정 코드만으로 제한 해서 검색하는 기능이다. 웹 페이지 검색의 경우 호스트나 사용자별 최상위 문서만을 대상으로 검색하는 기능 등으로 응용할 수 있는 기능을 제공한다.

4) 애니서치

한글, 한자, 외래어 등 언어의 특성을 고려한 인공지능 색인(애니색인)이 가능하고 검색결과는 가중치를 적용하여 적합성 순으로 해당 문서에 대한 요약물 첨가 해서 출력해 준다.

(1) 부울연산자

기본적인 부울연산자를 사용하고 있다. 즉 단어를 모두 포함하는 그리고, AND, & 와 입력된 검색어 중 하나만을 포함해도 검색하는 또는, OR, + 와 , 입력한 단어가 들어가지 않도록 하는 NOT와 !을 사용해 검색식을 완성할 수 있다.

(2) 절단검색이 가능하므로 정보* 로 검색하면 “정보”로 시작하는 모든 문서를 검색해 준다.

(3) 연산자 순서는 NOT AND OR 순이며 우선순위 변경시에는 “()”를 사용

한다.

(4) 검색의 신속성을 위해 한글자는 검색하지 않는다.

5) 유니파인더

검색한 정보에 대해서 요약문을 실어 자체순위결정 알고리즘을 통하여 가중치를 적용해 10개씩 출력한다. 한글, 영어, 한자 검색이 가능하고 영어로 된 검색어가 국내 웹페이지 뿐만 아니라 국외 문서들도 검색한다. 그러나 국외 사이트의 문서는 아직 부족한지 검색되는 양이 적다. 부울연산자 사용, 절단검색, 연산순서는 "에니서치"와 동일한 방법을 사용한다.

6) 웹글라이더

대부분의 검색엔진들이 상용온라인 서비스에 비해 검색옵션이 다양하지 못한 면이 있다. 이러한 간단한 검색옵션들은 초보이용자에게는 사용에 편리함을 주지만 정확한 결과를 원하는 사람이나 검색에 익숙한 고급 전문가에게는 아쉬움이 되어왔다. 그러한 면에서 국내 웹페이지를 대상으로 하는 상용 검색엔진 중 처음으로 위치연산자를 지원하다는 면에서 의의가 있다고 할 수 있을 것이다. 위치연산자는 검색어가 속해 있는 문서 뿐만 아니라, 검색어의 위치정보를 추적함으로써 검색어의 순서와 위치를 사용자가 결정, 원하는 정보에 보다 정확하게 접근하게 해준다. 또한 깊이 검색기능을 활용하면 웹사이트내 하부 디렉토리 홈페이지로 검색범위를 좁힐 수가 있다.

(1) 부울연산자

기본적인 부울연산자를 사용하고 있다. 즉 AND, OR의 역할을 하는 &와 스페이스, +와, 그리고 NOT의 역할을 하는 !을 사용해 검색식을 완성할 수 있다

(2) 위치연산자

위치검색은 사용자가 입력한 키워드들 사이의 거리를 정하여, 해당 거리내에 키워드들이 존재하는 문서를 검색하는 기능으로 검색한 결과의 양이 많을 때 특히 유용하게 이용할 수 있다. 예를 들면 정보~2 검색으로 검색하면 '정보'와 '검색'이 순서에 상관없이 2단어 이내로 붙어 있는 문서를 검색하고 정보^2 검색으로 검색하면 '정보'와 '검색'이 순서대로 2단어 이내로 붙어 있는 문서를 검색한다.

(3) 문장검색

문장검색은 사용자가 입력한 키워드들이 한 문장내에 존재하는 문서를 검색하는 기능이다. 예를 들면 다음과 같다. @(정보 검색)으로 검색하면 '정보'와 '검색'이 한 문장내에 있는 문서를 검색한다.

(4) 도메인검색

국내 웹사이트를 대상으로 하는 것이 아니라 사용자가 원하는 도메인내에서만 검색을 실행한다.

(5) 깊이검색

웹문서가 존재하는 디렉토리의 깊이를 사용자가 정하는 것으로 도메인 검색과 함께 방대하게 검색되는 문서들을 줄이는 방법으로 좀더 사용자가 원하는 문서에 접근할 수 있다. 깊이 0을 선택하면 국내

웹사이트의 홈페이지 내에서만 검색을 수행하고 깊이 1을 선택하면 홈페이지 다음의 서브디렉토리까지를 깊이 2를 선택하면 홈페이지 다음의 서브디렉토리 2개까지를 보여준다.

(6) 구문검색

여러 단어가 결합하여 사용된 인용구를 효과적으로 검색할 수 있다. 주로 영화제목, 가요명, 도서명이 인용구로 사용된다.

(7) 절단검색

정보* 로 검색하면 “정보”로 시작하는 모든 문서를 검색해 주며 *정보로 검색하면 “정보”로 끝나는 모든 문서를 검색해 준다.

(8) 필드검색

제목과 본문을 지정해서 해당되는 필드에서만 검색해 준다.

웹글라이더에서 지원하고 있는 메타글라이더의 검색방법에 대해 간단히 부연 설명하면 다음과 같다.

검색어를 입력할 때는 단어를 스페이스로 구분하고 있다. 모든 검색엔진과 원하는 검색엔진을 선택하여 사용할 수 있고 검색된 결과 형태는 본문과 요약 가운데 선택할 수 있고 검색시간도 조절할 수 있다.

1. AND, OR 연산자를 사용하며 단어 사이는 띄어쓰기를 한다.

2. 두단어 이상의 단어에 대해 구문검색을 옵션으로 지정할 수 있다.

3. 2. 3 검색엔진의 정보검색 옵션 비교분석

이상은 국내 주요 검색엔진 서비스의 주요 검색 기능 이용법이다. 해외 검색엔진 서비스들은 영문으로 된 자료가 적은 국내 사이트들의 문서에 대한 배려가 적을 수 밖에 없고 대부분의 해외 검색엔진 서비스들은 한글로 된 검색 단어를 이용할 수 없다. 이러한 상황에서 국내의 웹문서들을 대상으로 검색할 수 있을 뿐 아니라 한글의 고유한 특성을 이용해서 편리하게 검색해 주는 검색엔진 서비스들이 개발되고 있다는 것은 이용자와 웹사이트 운영자에게 모두 바람직한 것이라고 생각된다.

그러나 국외 검색엔진들에 비해 인접연산처리나 확장검색이 미흡하다. 또한 같은 검색 기능이라도 각 검색엔진마다 서로 사용하는 표기법이 달라서 이용자에게 혼란을 초래하기도 한다. 국외 검색엔진들은 검색 단어들을 입력하는 방법들이 서로 유사한데 비해서 국내 검색엔진들은 특히 기본적인 부울연산자 처리에서 그렇지 못하다. 부울연산자의 경우는 대부분의 검색엔진들이 대부분 지원하고 있으며, 초보 이용자들도 쉽게 사용하는 것인데 적어도 이에 대한 표준화만 이뤄진다고 해도 이용자에게는 대단히 편리한 옵션들이 될 것이라고 생각한다. 또한 공백처리에 있어서도 어떤 검색엔진은 “또는”으로 처리하는가 하면 어떤 검색엔진은 “그리고”로 처리를 하고 있다. 쉽게 자주 사용되는 검색기능에 있어서 이용자의 인터페이스를 고려한 표준화가 필요하다고 본다. 검색하고자 하는 것이 아주 특정적이거나 뚜렷

한 것이라면 각 검색엔진들의 특별한 사용법을 잘 알지 못해도 상관없을지 모르지만 검색된 결과가 너무 많다면 검색옵션을 사용해서 제한해 주면 훨씬 더 정확한 정보를 검색해 낼 수 있다.

검색어에 따른 검색결과에 대해 컴퓨터가 주는 신뢰점이라고 할 수 있는 적합성의 정도를 표시해 주는 가중치 표시는 대부분의 검색엔진들이 가중치를 표시해 주고 있는데 높을수록 정확한 검색결과임을 뜻한다고 할 수 있다. 그러나 절대적으로 신뢰할 수 있는 것만은 아니다. 실제로 상대적으로 낮은 가중치를 보인다고 해도 검색자가 원하는 정보일 수도 있기 때문이다. 검색도구의 성능을 평가하는 많은 논문들이 이러한 컴퓨터가 부여한 가중치순에 따라 위로 부터 10개의 논문들을 가지고 적합성을 평가하는 경우가 있다 (Chu, H. and Rosenthal, M., 1996). 그러나 웹문서들이 방대한 만큼 검색해주는 문서들도 많으며 가중치란 대체로 검색어로 판정되는 것이기에 문서의 내용을 대표한다고는 볼 수 없으므로 top 10개만의 문서들로 검색도구의 정확율을 평가하기에는 무리라는 판단이 된다. 뿐만 아니라 국외의 검색엔진들과 마찬가지로 국내의 경우에도 가중치를 부여하는 방법들이 서로 달라서 제목이나 요약정보를 기준으로 검색어에 대한 가중치를 부여하는가 하면 어떤 것은 본문까지도 포함되기도 하는데 무엇보다도 본문을 기준으로 하는 가중치 부여가 필요하며 이에 대한 도움말 기능도 필요하다고 본다. 마지막으

로 검색엔진에 따라서는 같은 검색어를 입력해도 서로 아주 상이하거나 때로는 의도하지 않는 정보를 찾을 때가 있으므로 검색어 선정에 주의를 기울여야 원하는 정보를 찾을 수 있다.

3. 3 검색도구의 활용방안

본 결과들을 정리해 보면 <표 1>, 우선 심마니는 검색어 오류확장, 한영확장, 발음확장 등의 기능을 수행한다. 기호 뿐만 아니라 자연어로된 부울연산자 처리가 가능해서 연산자 사용에 익숙하지 않은 초보자나 검색하고자 하는데 있어서 검색어를 확실히 알고 있지 못할 경우에 더욱 유용하다. 뿐만 아니라 한번의 검색어 입력으로 분류항목, 사이트, 웹문서, 신문기사, SEARCH.COM, Yahoo, AltaVista 까지 검색해서 출력하여 주므로 그 어떤 검색도구보다도 다양한 검색결과들을 볼 수가 있다. 뿐만 아니라 최근에는 국외의 cnet에 대한 요약정보와 함께 필요한 정보에 대해 직접 링크할 수 있도록 하고 있어 컴퓨터 분야에 관심있는 이용자에게 좋은 서비스를 제공하고 있다. 한편 심마니가 지원하지 않는 뉴스그룹에 대한 검색은 까치네, 정보탐정, 유니파인더, 애니서치 등을 이용하면 좋는데 무엇보다도 삼성종합기술원의 뉴스서버를 이용하고 있는 유니파인더가 유즈넷을 검색하기에 좋다. 뿐만 아니라 한글과 영문을 각 언어의 특성에 맞는 방법으로 색인하므로 영문을 지원하는 다른 검색도구보다 정확한

정보를 찾기에 좋다. 그리고 새롭게 단장된 정보탐정Ⅱ에서는 유즈넷검색 뿐만 아니라 한영성경검색을 추가로 서비스 하고 있으므로 다른 검색도구에서는 지원하지 않는 성경검색에 유용하다.

그리고 하나의 검색어만으로 여러 신문의 기사를 검색해 보고자 한다면 미스 다 찾니나 애니서치를 이용하는 것이 좋다. 신문통합검색기능을 가지고 있으므로 각 신문기사들의 고유한 검색방법을 모르거나 시간을 절약하고 싶을 때 사용하면 좋다. 부울연산자는 물론이고 위치연산자, 구문, 절단, 필드, 도메인, 깊이검색 등 다양한 검색옵션이 있어 정확한 검색결과를 얻어낼 수 있는 검색도구는 웹글라이더로써 검색옵션 사용에 익숙한 고급 검색자에게 특히 유용하다. 특히 검색결과가 너무 많을 때 다양한 옵션으로 검색하여 정확한 결과를 산출해 낼 수 있다.

한편 각 검색도구들의 URL을 모르거나 검색도구 선정에 어려움을 느낀다면 메타검색엔진인 동시에 통합검색기능을 제공하는 미스 다찾니를 이용하는 것이 좋다. 각 검색도구들만의 고유한 검색옵션을 사용할 수 없는 단점이 있지만 국내외 검색도구 및 국내외 신문기사와 그밖의 여러 다른 사이트들이 링크되어 있어서 다른 곳으로 이동하지 않고도 한 곳에서 다양한 검색을 할 수 있다.

그러나 검색어 선정에 어려움이 많다면 목록서비스를 이용해 보는 것이 좋다. 위에서 설명한 검색엔진들과는 달리 목록서비스는 검색식 작성이 필요하지 않아 정

보검색에 편리할 뿐만 아니라 특정 주제별로 각 페이지를 분류하여 정리해 놓음으로써, 유사한 주제 정보를 알리는 사이트를 한 눈에 볼 수 있도록 하고 있다. 목록서비스 가운데 특히 DIR은 같은 목록서비스를 제공하고 있는 다른 곳에 비해서 사용자 인터페이스가 좋고 분류갯수도 많아 검색결과의 양도 많다. 뿐만 아니라 국내 웹관련 뉴스게시판에 대한 서비스를 통해 웹에 대한 상세한 정보를 얻기에 좋다. 이곳에서 제공하는 추천사이트는 넷퀘스트21에서 추천하는 사이트와 같이 단순한 브라우징에도 좋다.

이렇듯 목록서비스는 검색엔진과는 달리 주제별 정보에 대한 사이트를 제공해주므로 유사한 주제에 대한 다양한 사이트를 알 수 있고, 주제분류체계만 잘 익히면 어떠한 정보라도 같은 방법으로 찾을 수가 있다. 그러나 실제로 접속해 보기 전에는 그 사이트에서 제공하는 정확한 정보를 찾기가 어렵고 그 사이트의 이름과 함께 등록된 간략한 설명을 통해서만 사이트를 선택해야 하므로 많은 시행착오를 겪게 된다. 또한 원하는 정보를 제공하는 사이트를 찾더라도 원하는 정보를 보기 위해서는 다시 한 번 해당 사이트로 접속해야 한다는 단점이 있다(W3 제 5호). 그러므로 정보요구에 따라 웹문서상의 정보가 필요하다면 검색엔진을, 주제별 정보를 담고 있는 사이트가 필요하다면 목록서비스를 이용하는 것이 좋으며, 때로는 두가지를 동시에 사용해야 만족스러운 결과를 얻을 수도 있다. 각 검색도구의 장단

점들을 <표 3>에서 보여주고 있다.

이와 같이 목록서비스와 검색엔진이 갖는 각각의 장단점을 보완하기 위해서 근래에는 많은 검색도구들이 검색엔진과 목록서비스를 동시에 병행하거나 목록서비스 상에서 키워드 검색이 가능하도록 하는 추세이다. 국내의 심마니가 전자의 대표적인 예이고, 국외의 유명 목록서비스인 Yahoo가 후자의 대표적인 예이다. 국내에서도 많은 검색도구들이 전자와 같은 방식으로 두가지 방식을 모두 채택하고 있다 <표 3>. 그러나 전자의 경우에 있어서 심마니를 제외하고 대부분이 키워드 검색으로는 문서만을 검색할 뿐 주제별 사이트는 검색해 내지 못하고 있다. 또한 후자의 경우에 있어서도 키워드 검색이 현재의 주제범위에서만 가능한 것이 있는가 하면 현재 위치하고 있는 서브 카테고리에 상관없이 전주제에 걸쳐 키워드 검색이 되는 경우가 있어 이에 대한 표준화내지는 자세한 도움말 기능이 필요하다고 본다. 그러나 이용자의 입장에서 볼 때, 잘못된 주제 카테고리 들어갔을 경우 원하는 정보를 찾을 수가 없는 목록서비스 상의 단점을 생각할 때, 전주제에 걸쳐 키워드 검색이 가능한 서비스가 이용자에게 더욱 유효하리라고 보여진다. 끝으로 <표 2>에서 보는 바와 같이 대부분의 검색도구들은 검색해 낸 문서나 온라인 사이트에 대해 요약정보를 제공해 주는데 이것만으로 적합성을 판단하기가 어렵다면 검색도구에서 평가를 제공해 주는 ZIP나 인사이드 코리아를 이용해 보는 것도 좋

다. 그러나 인사이드 코리아는 검색되는 양이 적으므로 재현율이 높은 것을 원할 때는 적합하지 않다.

4. 결론

본 연구에서는 국내에서 개발되어 가장 많이 활용되고 있는 검색도구를 목록서비스와 검색엔진으로 구분하여 그 특성을 살펴 보았다. 앞에서 살펴본 바대로 각각의 검색도구들은 나름대로의 특성과 장단점<표 3>을 가지고 있기 때문에 그것을 비교하고 평가하는 일은 쉽지 않다. 그러므로 선행된 국외 검색도구들의 검색효율과 성능을 평가하는 방법들이 연구마다 서로 다양한 것이다.

그러나 이러한 기준들 외에 검색도구들의 로봇 에이전트 유무, 질의연산능력, 갱신주기, 부가적인 서비스 등도 중요시되므로 <표 1, 2>에서 이러한 점들을 기준으로 정리해 보았다. 그러나 검색도구의 성능을 평가하기 위해 무엇보다도 중요시 되는 데이터량과 갱신주기에 관해 밝히고 있는 검색도구가 적고 이에 대한 자료도 적어 자세한 조사가 이루어지지 못했음을 밝힌다. 이는 아마도 국외 검색도구에 비해 자체 데이터의 양이 현격히 적고 이에 대해 밝히기를 꺼려하는 것 같다. 그러나 이와 같은 기준들은 검색도구를 평가하는데 있어 무엇보다도 중요한 자료가 되므로 각 검색도구의 운영자들은 이를 밝히는데 있어 주저하지 말아야 할 것이다.

아직까지 모든 서비스를 제공해 줄 수 있는 검색도구는 없다. 여러 검색도구를 사용해서 각각의 장점을 익혀 활용할 필요가 있다. 따라서 찾고자 하는 정보의 성격에 따라 검색도구를 잘 선정해야 한다. 또한 검색도구마다 각각의 특성과 검색수행능력은 모두 다양하므로 검색을 시작하기 전에 지시사항이나 도움말 기능, 검색방법, 데이터베이스의 구성방법 등에 대해 브라우징 할 필요가 있다.

국내 검색도구의 설계와 개발은 국외의 검색도구들에 비해 늦게 시작되었고 따라서 보유하고 있는 데이터의 량이나 분류 목록의 갯수가 국외의 유명검색엔진들에 비해 현격히 적고 부가적인 서비스 기능

들 또한 아직은 미흡한 수준이다. 그러나 한글의 고유한 특성을 파악해서 국내의 유용한 웹문서들을 이용할 수 있도록 해 준다는 면에서 국내 검색도구의 의의를 찾을 수 있을 것이다. 웹의 유효성에 따라 국외의 검색도구들에 대한 다각적인 평가 방법들이 제안되고 있는데 국내 검색도구에 대한 평가방법들은 아직 전무한 상태이다. 인터넷의 이용이 더 이상 특정인을 대상으로하는 것이 아닌 일반인을 대상으로 확대되어 나간다는 것을 감안할 때, 한글을 이용할 수 있는 유용한 국내 검색도구의 설계 및 개발과 이에 대한 적절한 평가방법이 무엇보다 중요하다고 보여진다.

〈표 1〉 검색엔진의 정보검색 옵션 비교

	부울연산자 AND OR NOT	구절검색	연산순서	인접연산 또는 위치연산	절단검색	필드제한 검색	유의어 집단검색 확장검색	가중치	기타
까치네	+ 공백처리 NOT					본문 도메인		적용	
애니서치	그리고,& AND, 또는,OR, +NOT, !		NOT,AN D,OR순이 며 순서를 바꾸고 싶 을 때는 () 를 사용 한다.		*			적용	
웹글라이더	&,공백처 리 + , !	" "		~, ^	*	본문제목	확장검색 (옵션에서 지정)	적용	문장검색 깊이검색
유니파인더	그리고, AND, & 또는, + OR NOT		애니서치 와 같다.		*			적용	대소문자 구별
정보탐정	AND,*,& OR,+ -	" "		{ }	%		확장검색	적용	대소문자 구별
십마니	그리고, 와,과,및, &,AND +,또는, OR,혹은 !&, &!	" "	()		*	도메인	유의어 한글확장 영어확장 발음확장 #	적용	대소문자 구별

〈표 2〉 검색도구에 대한 비교

	데이터량	로봇 에이전트	검색대상	갱신주기	웹페이지	요약정보	쪽당표시
까치네	50만이상	유	국내웹 유즈넷	?	전체	본문내용 가운데 요약하고 검색어는 하이라이트로 표시	10
애니서치	?	유	국내외웹 유즈넷신문	?	?	사이트에 대한 간단한 소개	10
웹글라이더	?	유	국내웹	?	?	검색어에 따라 요약정보가 달라지며 검색어는 하이라이트로 표시	10
유니파인더	30만 이상	유	국내외웹 유즈넷 신문	?	?	본문 내용 가운데 요약	10
정보탐정	22만~30만	유	국내웹과 국내웹과 연결된 국외웹 유즈넷	?	전체	사이트에 대한 간단한 소개	10
심마니	약38000분 류목록개수 12811	유	국내웹	2주일	전체	검색어에 맞는 문서 내용을 요약하고 검색어는 하이라이트로 표시	10
코식크	디렉토리 서비스가 가능한 홈 페이지 1200개 이 상(지속적 인 추가중)	?	국내웹	?	?	간단한 소개	없음
DIR	분류목록개 수?	사람에 의한 주제분류	국내웹관련 뉴스 게시판	?	?	디렉토리외 각 서브디렉토리 및 간단한 요약	없음
ZIP	?	사람에 의한 주제분류	대부분이 국내웹	?	?	검색어에 따라 간단한 소개와 문서요약 등 다양	없음
넷퀘스트	?	?	국내웹	?	?	간단한 소개	없음
미스다찾니	없음	인터페이스 에이전트와 검색에이전 트	국내외웹신 문	관련없음	관련없음	문서요약이나 간단한 소개와 찾은 정보의 출처 표시	없음

〈표 3〉 검색도구의 장단점

검색도구 종류	장 점	단 점
코식크	영문지원 메타서치 가능	검색속도가 다소 느림 연산자 기능이 미약
DIR	체계적인 주제분류와 그에 따른 검색결과도 좋음	
ZIP	주제별 디렉토리 내용을 검색어로도 검색이 가능	
넷웨스트21	영문지원 메타서치 가능	도움말 기능 미약검색되는 정보의 양이 다 른 것과 비교해 상당히 적음 심마니
심마니	메타서치 가능 주제별 분류서비스의 추가적인 서비스 자연어로 된 부울연산자 처리가 가능 이용자 인터페이스를 고려 한글 외래어를 영문으로 변환하여 검색 한번의 검색으로 사이트, 웹문서, SEAR- CH COM, Yahoo, Alta Vista까지 검색 cnet 등 추가적인 다양한 서비스 제공	영문을 지원하지 않음 유즈넷 검색을 지원하지 않음 홈페이지가 다소 복잡함
까치네	메타서치 주제별 분류서비스의 추가적인 서비스 데이터량이 현재 가장 많다 충실한 도움말 기능과 검색엔진 설계자를 위한 검색엔진 원리에 대한 설명이 우수	검색속도가 다소 떨어짐 검색어 입력으로는 분류서비스를 이용하지 못함
정보탐정	메타서치 주제별 분류서비스의 추가적인 서비스 데이터량이 현재 가장 많다 충실한 도움말 기능과 검색엔진 설계자를 위한 검색엔진 원리에 대한 설명이 우수 고속색인 및 검색으로 대용량의 문서 수용 전화번호 이용에 좋음	
애니서치	메타서치, 영문지원 체계적인 주제분류	
유니파인더	메타서치 가능 유즈넷과 신문검색에 유리	
웹글라이더	메타서치 다양한 연산자 지원 메타글라이더로 메타검색엔진 서비스	부가적인 서비스 기능이 다소 약함

참고문헌

- 이란주, 1997. "월드 와이드 웹(World Wide Web) 검색엔진의 효율적인 검색방법 고찰 - Lycos, InfoSeek, Alta Vista, Excite, Yahoo 를 중심으로." 도서관. 52(2) : 29-43.
- 정영미, 김성은, 1997. "WWW 탐색도구의 색인 및 탐색 기능 평가에 관한 연구", 한국문헌정보학회지. 31(1) : 153-184.
- 마이크로소프트웨어 96년 4월호.
월간 뉴미디어 96년 6월호.
컴퓨터 월드 96년 4월호.
한글판 PC WEEK, 97년 3월호.
경향신문 97년 2월 26일.
동아일보 97년 2월 27일.
전자신문 97년 2월 27일, 3월 1일.
정보통신신문 97년 3월 3일.
중앙일보 97년 2월 27일, 3월 1일, 3월 10일.
매일경제신문 97년 2월 26일, 3월 1일.
문화일보 97년 3월 4일.
한국경제신문 97년 2월 27일, 97년 3월 1일.
MYCOM 96년 5월호.
HOME PC 96년 4월호.
PC LINE 96년 4월호.
PC PLUS 96년 6월호.
PC WORLD 96년 4월호.
W3 제5호 제6호.
Belkin, N. C. et al. "The Effect of Multiple Query Representations on Information Retrieval System Performance," on Proceedings of the Sixteenth Annual International ACM-SIGIR Conference.
- Chu, H and M. Rosenthal, 1996. "Search Engines for the World Wide Web : A Comparative Study and Evaluation Methodology." Presented at the 96' ASIS Conference.
[http://www.asis.org? annual-96/ElectronicProceedings/chu.html](http://www.asis.org?annual-96/ElectronicProceedings/chu.html).
- Courtois, M. P. et al. 1995. "Cool Tools for Searching the Web : A Performance Evaluation." Online. 19(6) : 14-32.
- Kimmel, S. 1996. "Robot-generated Databases on the World Wide Web" Database. 19(1) : 40-49.
- Krol, E. 1993. "The Whole Internet : User's Guide & Catalog. Sebastopol". O'Reilly & Association, Inc.
- Lancaster. F.W. and E. G. Fayen. 1973. Information Retrieval. On-Line Los. Angeles, CA :

- Melvill Publishing Co.
- Leighton, H. V. 1995. "Performance of four World Wide Web (WWW) Index Service : Infoseek, Lycos, Web-Crawler, and WWWorm." <http://www.winona.msus.edu/services-f/library-f/webind.htm>.
- Leonard, A. J., 1996. "Where to Find Anything on the Net." <http://www.cnet.com/Cotent/Reviews/Search>.
- Marchionono, G. and D. Barlow. A Comprision of Boolean-based Retrieval to the WAIS System for Retrieval of Aeronautical Information : Final Report.
- Scoville, R. 1996. "Find it on the net!" : Special Report. PC World. <http://www.lycos.com>.
- Sullivan, D. 1997. "A Webmaster's Guide to Search Engines and Directories." <http://calafia.com.webmasters/>.
- Zorn, P. et al. 1996. "Advanced Web Searching : Tricks of the Trade." Online 20(3) : 14-28. [http://bandi.cs.pusan.ac.kr/irhelp.htm-2671 bytes](http://bandi.cs.pusan.ac.kr/irhelp.htm-2671bytes). <http://namu.museum.go.kr/ir-query-syntax.html>. <http://selfare.or.kr/search.htm>. <http://fact.kordic.ac.kr/kosti96/proceeding/d2>. <http://daelim.co.kr/lecture>.