Historia Mathematica Vol. 12, No. 1, 1999, 82-87

통계관련 사이트의 검색엔진에 관한 고찰

서경대학교 응용통계학과 **이승우**

Abstract

The Internet has developed the computer and communications world. The Internet has tons of millions of sites. To find the specific information, we use a search engine. All search engines are operated by keyword against a database, but many different factors affect the result of search by engines. In this paper, we investigate the development of the Internet and try to find the differences among the most popular search engines.

0. 인터넷의 탄생과 발전

1960년대 중반, 미·소간의 냉전 시대에 미국방성은 소련의 선제 핵공격에 의하여 핵무기를 제어할 수 있는 컴퓨터가 파괴되더라도 다른 컴퓨터에서 제어할 수 있도록 하는 컴퓨터 네트워크를 구상한 것이 인터넷의 시초이다. 이에 따라 국방부 산하 고등기술연구소에서 유타 대학, 샌타바버라의 캘리포니아 대학, 로스앤젤레스의 캘리포니아 대학, 스탠퍼드 대학의 4대의 대형 컴퓨터를 연결하여 NCP라는 프로토콜을 기반으로 알파넷(ARPnet)을 1969년에 탄생시켰다. 알파넷의 기능 향상과 더불어 사용자의 증가로 인하여 다양한 컴퓨터의 기종에 서로 다른 통신망을 연결할 수 있는 적합한 TCP/IP라는 새로운 프로토콜을 개발하였으며 1983년에 완전히 TCP/IP로 전환되었다. 또한 ARPnet은 1980년대 초, 군사 부분 통신망 MILnet과 비 군사 부분 통신망-민간 연구 통신망인 ARPnet인 두 개의 네트워크로 나뉘어지게 되었다. 1986년에 미국립과학재단에서 슈퍼컴퓨터간의 NSFnet이라는 통신망을 운영하여 대학, 연구기관, 학술 단체, 기업들이 사용함으로서 그 규모가 급속도로 확장되었으며 세계 각국의 통신망도 급속한 속도로 NSFnet에 연결되어 세계의 통신망을 연결하는 인터넷이 탄생되었다. 인터넷은 1990년 이후 월드 와이드 웹 서비스가 시작되면서 이용자도 일반화되고 있으며 현재 사용자가 증가 추세에 있다. 그러므로 본 논문에서는 대표적인 3가지 검색엔진을 이용하여 인터넷에서 제공되는 통계학의 다양한 정보를 고찰하고자 한다.

1. 주제별 검색엔진

여러 분야의 각 사이트를 특정한 주제별로 분류해서 원하는 정보를 찾을 수 있는 검색엔진을 말한다. 이 검색엔진의 장점으로는 정보에 대한 사전지식이나 주소를 알지 못해도 찾을 정보의 분류를 따라가면서 쉽게 검색이 가능하며, 단점으로는 검색을 위한 많은 노력과 시간이 필요하며 한번 잘못된 분류를 따라가면 올바른 정보를 찾기 어렵다는 것이다.

Yahoo(http://www.yahoo.com)는 대표적인 검색엔진으로 1994년 스탠퍼드 대학교의 대학원생이였던 데이비드 필로와 제리 양에 의해서 만들어졌다. 특징으로서, 14개의 주제별로 계층적으로 사이트가 정리되어 있으며, 주제별 하위분류에서 주제어를 선택, 검색이 가능하기 때문에 정확한 검색을 할 수 있으며 주제별 검색과 단어별 검색을 모두 지원한다. Yahoo의 통계 관련 사이트 검색으로서 검색어를 statistics로 입력하면 자료를 찾을 수 있고, 또한 Home>Science>Mathematics>Statistics>의 순서로 검색이 가능하다. 검색어가 applied probability theory인 경우 Math Pages의 http://www.seanet.com/~ksbrown/iprobabi.htm, 검색어가 sampling theory인 경우 http://www.dsig.com/methods/foutline.html, 검색어가 decion theory인 경우 Bayesian Links http://psych.fullerton.edu/mbirnbaum/bayes/bayeslinks.htm, 검색어가 design and analysis of experiment인 경우 http://www.statisticaldesigns.com/, 검색어가 multivariate analysis인 경우, http://www.camo.no/Applications/MultivariateAnalysis.html

야후 코리아(http://www.yahoo.co.kr)는 1997년 9월 한글로 메뉴구성을 하여 서비스를 하고 있다. 야후 코리아의 통계 관련 사이트 검색은 처음>자연과학>수학>통계학>으로 국내 대학의 통계학과 홈페이지와 개인 홈페이지를 볼 수 있으며, 또한 처음>자연과학>수학>에서 유즈넷을 선택하면 han.sci.stat - 통계학 관련 뉴스그룹을 볼 수 있다. 검색어로 통계학을 입력해도 같은 결과를 얻을 수 있다. 처음>사회와 문화: 사람들: 개인 홈페이지>

심마니(http://www.simmani.com)는 한컴네트의 자연어 처리팀이 개발한 한글 검색엔진으로서 현재는 데이콤에서 운영하고 있으며 규모가 큰 데이터 베이스를 구축하고 있다. 특징으로서, 16개 주제로 분류되어 있으며 다양한 연산자를 지원하고 있다. 주제별 검색과 단어별 검색 모두를 지원하며, 한글 및 영어 연산자가 입력가능하며, 한글 유의어 사전의 참조기능, 한글 외래어의 영어 참조기능이 가능하며 신조어 인식 기능도 있다. 또한 찾은 정보들을 우선 순위별로 표현하는 장점이 있다. 심마니의 통계 관련 사이트 검색으로서, 검색어로 통계학을 입력하면 국내 대학의 통계학과 홈페이지가 링크 되어 있지만 원하는 자료를 찾기 어렵다. 통계를 검색어로 사용하면 통계학보다는 일반적인 통계 자료가 검색된다.

코시크(http://www.kor-seek.com)는 충남대학교에서 개발한 국내 최초의 웹 검색엔진으로 처음에는 한글검색엔진으로 시작, 현재는 영문서비스도 하고 있는 검색엔진으로 현재

는 웹 코리아로부터 후원을 받고 있다. 특징으로서, 12개 주제로 분류되어 있으며 주제어 및 주제별 검색, 확장 검색을 지원한다. 그러나 연산자가 지원되지 않아 검색이 불편하다. 주제어 검색시 단어는 최대 6개까지 입력이 가능하며, 정보의 일반적인 단어를 사용함으로써 검색 결과를 정확하게 확인 할 수 있다. 코시크의 통계 관련 사이트 검색으로서, 검색어로 통계학을 입력하면, 국내 대학의 통계학과 홈페이지가 링크 되어 있다. 처음>환경, 과학>수학으로도 검색이 가능하며 이곳에는 수학에 관련된 홈페이지와 개인 홈페이지가 링크 되어 있다. 수학이라는 검색어에서는 각 대학의 수학과 홈페이지가 검색된다.

한국디렉토리 움프(http://www.oomph.net)는 1995년 한국의 기업정보 및 국내의 모든 정보를 분류하여 인터넷상에서 한국을 대표하는 디렉토리로 출발하여, 현재 정보검색과 디렉토리 서비스를 한글, 영문으로 서비스하고 있는 검색엔진이다. 특징으로서, 17개 주제로 분류되어 있으며 한, 영 주제별 검색을 동시에 지원한다. 다양한 부가서비스를 제공하며, 매주 새롭게 갱신되는 새로운 사이트를 소개한다. 한국디렉토리 움프의 통계 관련 사이트 검색으로서, 검색어로 통계학을 입력하면 국내 대학의 통계학과 홈페이지가 링크 되어 있다. 검색어로 수학을 입력하면 수학과 홈페이지와 개인 홈페이지가 검색된다. 또한 여행사의 수학 여행 등이 검색된다. 처음>환경, 과학>수학으로 검색도 가능하다.

Galaxy(http://galaxy.tradewave.com)는 Tradewave에서 제공하는 검색 엔진이다. Galaxy의 디렉토리는 인터넷상에서 수천 개의 사이트에 연결되어 있다. 특징으로서, 다양한 방법(Link text, Title text, All text)으로 검색이 가능하며 Gopher의 제목이나 Telnet의 자원을 검색할 수 있다. Galaxy의 통계 관련 사이트 검색으로서, 검색어로 statistics, mathematics를 입력하면 국외 대학의 통계학과 관련 홈페이지가 링크 되어 있다. Science>Mathematics>Statistics>로 검색이 가능하다. 또한, Advanced Search에서 검색어를 statistics로 검색 가능하며 Science>Mathematics>Statistics>로 가면 sci.stat.math 같은 통계관련 Discussion Group을 볼 수 있다.

Magellan(http://magellan.mckinley.com)은 미국의 McKinley그룹에서 운영하고 있는 검색엔진으로서 단어별 검색도 가능하다. 특징으로서, 18개 주제로 분류하며 주제별 검색과 단어별 검색을 지원한다. 영어, 불어, 독어를 지원하며 수시로 Update하면서 내용의 깊이, 탐색의 용이성과 디자인의 독창성, 참신함 등에 의한 자체적인 기준으로 평가, 별표로 점수를 부여한다. Magellan의 통계 관련 사이트 검색으로서, 검색어로 statistics, mathematics를 입력하면 국외 대학의 통계학과 관련 홈페이지가 링크 되어 있다.

2. 단어별 검색엔진

수많은 사이트를 하나의 데이터 베이스화 한 후, 사용자의 검색어의 입력을 통해 미리 수집된 자료 중에서 해당 조건에 맞는 사이트를 찾아주는 검색엔진을 말한다. 이 검색엔진의

장점으로는 검색어로 원하는 정보를 쉽고, 빠르게 찾을 수 있으며, 단점으로는 검색어가 정확하지 않으면 정보를 찾기 어렵다. 해당 검색엔진이 많은 자료를 데이터 베이스에 정확하게 보유하지 않은 경우 원하는 정보를 찾지 못할 수도 있으므로 단어의 선정이 중요하다.

Altavista(http://www.altavista.com)는 Digital Equipment Corporation사가 1995년 팔로 알토에 의해 개발되었으며 알파기술을 이용하여 검색할 수 있다. 특징으로서, 방대한 자료로부터 단순검색, 조건검색이 가능하다. 25개국 언어를 지원하며, 유즈넷 검색도 가능하고 검색어와 날짜별로도 검색이 가능하다. 다른 검색엔진과는 달리 한글, 일본어, 한자 등 2바이트로 구성된 단어도 검색이 가능하다. 통계 관련 사이트 검색으로서, 검색어로 statistics, mathematics를 입력하면 국외 대학의 통계학과 관련 홈페이지가 링크 되어 있다.

한글 알타비스타 (http://www.altavista.co.kr)는 수식어(+, −)를 사용하여 검색이 가능하며 특정 혹은 전문화된 검색이 필요할 경우 조건 검색 기능을 사용하여 검색한다.

정보탐정(http://www.infocop.com)은 1996년 지능형 하이텔 안내시스템의 개발을 위하여 제작되었으며, 그해 6월 윙윙윙이라는 이름으로 서비스를 시작하였다. 1997년 6월 정보탐정으로, 1998년 9월 정보탐정 InfoCop으로 변경하였다. 현재 한국통신 멀티미디어 연구소 컨텐츠 서비스 연구팀에서 개발, 운영 중에 있으며 정보탐정, 인포캅, InfoCop은 같은 이름이다. 특징으로서, 영어단어 검색시 대소문자를 구별하며 다양한 연산자를 활용한 정밀한 검색이 가능하다. 어구 탐색을 이용하면 문서의 시작과 끝에 특정단어가 위치한 문서를 검색할 수 있다. 또한 다양한 부가 서비스(일본Web여행, 유즈넷서비스, 마이 뉴스, 멀티미디어서핑 등)를 제공한다. 정보탐정의 통계 관련 사이트 검색으로서, 검색어로 통계학을 입력하면 국내 대학의 통계학과 홈 페이지와 개인 홈페이지가 링크 되어 있다. 과학과 학문>자연 과학>수학>통계학으로 찾아갈 수도 있다. 검색어로 수학을 입력하면 국내 대학의 수학과 홈페이지와 개인 홈페이지가 링크 되어 있으며, 과학과 학문>자연 과학>수학으로 찾아갈 수 있다.

까치네(http://www.kachi.com)는 1996년 대구대학교에서 개발되었다. 특징으로서, 14가지의 주제로 분류, 서비스를 제공하며 다양한 검색연산자를 지원하며, 총 검색 결과를 숫자로 표시해준다. 검색내 재 검색 기능을 제공하며, 검색결과를 요약한 리스트를 제공한다. 까치네의 통계 관련 사이트 검색으로서, 검색어로 통계학을 입력하면, 국내 대학의 통계학과가 링크되어 있지만 제목이 불분명하고, 같은 홈페이지의 여러 페이지를 여러 번 검색하는 단점이 있다. 검색어 수학 역시 비슷한 결과를 얻는다.

Excite(http://www.excite.com)는 Excite사가 운영하며 주제별로 계층적으로 분류해서 주제별 검색도 지원하며 검색어의 개념비교를 통하여 고급의 정보검색이 가능하다. 따라서 검색어와 정확히 일치하지는 않지만 비슷한 뜻을 가진 정보를 검색할 때 좋은 검색엔진이다. 특징으로서, 약 5천만개의 웹 페이지를 검색어로 검색이 가능하고 약 6만개의 웹사이트를 주제별로 분류되어 있으며 최근의 유즈넷 뉴스들을 데이터 베이스에 저장하여 검색을 지

원하며 단어 검색시에 개념검색을 지원한다. Excite의 통계 관련 사이트 검색으로서, 검색어 statistics, mathematics로 검색시 대학의 홈페이지와 통계 관련 사이트를 검색할 수 있다.

InfoSeek(http://www.infoseek.com)은 InfoSeek사가 만든 상업용 단어별 검색엔진으로 www, Usenet, Ftp, Gopher등을 지원하지만 상업용이라서 비가입자는 www, Usenet의 뉴스그룹만 사용할 수 있는 제약이 있다. 검색 결과는 100개까지만 지원한다. 특징으로서 18개의 주제별로 분류되어 있으며, www와 Usenet의 뉴스그룹을 검색 할 수 있으며, 다양한 상업적인 기능을 가지고 있으며, 검색결과를 0~100까지의 정확도로 점수를 부여한다. InfoSeek의 통계 관련 사이트 검색으로서, 검색어 statistics로 검색시 통계 관련 사이트를 검색할 수 있다. Science>Mathematics>Statistics로도 검색이 가능하다.

Lycos(http://www.lycos.com)는 1995년 미국 카네기 멜론 대학의 Michael L. Mauldin에 의해서 개발된 검색엔진으로 처음에는 이미지와 사운드 파일만을 검색했으나 최근에는 웹 문서까지 검색한다. 또 Lycos는 문서에 포함된 이미지와 사운드 파일 등의 멀티미디어 검색을 할 수 있다. 특징으로서 18개의 주제로 분류되어 있으며 www의 대부분의 문서검색이 가능하며 Ftp, Gopher도 검색이 가능하다. 검색결과는 홈페이지 제목, 내용 등이 자세히 검색되며, 다양한 검색연산자 사용이 가능하며, 주제별로 검색을 할 수 있다. 검색결과에 대한 자체적인 평가 점수를 부여한다. Lycos의 통계 관련 사이트 검색으로서, 검색어 statistics로 검색을 할 수 있다. 또한, Science>Math>Statistics에서 Newsgroups으로 가면 검색이 가능하며 sci.stat.consult와 sci.stat.math 같은 뉴스그룹에 관한 것도 검색된다.

3. 메타 검색엔진

자체적인 데이터 베이스는 없으나 여러 개의 검색엔진을 한 페이지에서 사용할 수 있는 검색엔진을 말한다. 이 검색엔진의 장점은 지정된 검색어를 통해 여러 가지의 검색엔진을 동작시켜 보다 많은 검색결과를 제공하며 각각의 결과를 비교할 수 있다. 단점은 자체 검색엔진에서 검색이 아닌 메타 검색엔진을 거쳐야 하므로 속도가 느리며 검색 편차가 심하다.

미스 다찾니(http://www.mochanni.com)의 검색범위는 웹사이트와 신문이며, 검색을 의뢰하는 웹사이트는 한글 알타비스타, 야후 코리아, 정보탐정, 심마니, 코시크, Excite, InfoSeek, Lycos 등이 있다. 신문검색은 국내신문과 전자신문 그리고 CNN, 워싱턴포스트 등 국외 신문방송사들에서 의뢰한다. 특징으로서, 하나의 검색어로 다른 검색엔진들을 동시에 검색해서, 결과물을 볼 수 있으며, 영문 서비스도 제공한다. 단순한 메뉴에 비해 검색 범위는 방대하지만 각 검색엔진 고유의 검색방법을 사용할 수 없으며, 정밀한 검색 또한 기대하기는 어렵다. 미스 다찾니의 통계 관련 사이트 검색으로서, 검색어로 통계학을 입력하고 검색대상, 연산자, 처리시간 등을 Default값으로 검색하면, 대부분 각 학교의 홈페이지와 개인 홈페이지를 찾을 수 있으며 정보탐정, 심마니, 야후 코리아 등 어디에서 찾은 결과인지를

알 수 있다. 수학을 검색어로 사용하면 수학능력시험에 관련된 결과를 얻을 수 있다.

Savvy Search(http://www.savvysearch.com)는 대표적인 검색엔진으로서 Yahoo, Altavista, Amazon, DejaNews, InfoSeek, Lycos, WebCrawler, HotBot등 200개 이상의 의뢰된 검색 엔진들을 5개씩 나누어 검색하며, 검색 결과가 적당치 않을 때는 다른 검색엔진을 통해 연속적으로 검색이 가능하다. 23개 언어로 된 메뉴를 사용할 수 있다. 특징으로서, 검색 결과가 적당치 않을 때는 다른 검색엔진을 통해 5개씩 연속적으로 검색이 가능하다. Savvy Search의 통계 관련 사이트 검색으로서, 검색어 statistics, mathematics로 검색시 각 대학의 홈페이지와 통계 관련 사이트를, Mining Co., Yahoo!, Go To, Surf Point, LookSmart 등 5개의 검색엔진으로 검색하고 search more engines...을 선택하면 계속 다른 5개 검색엔진으로 검색할 수 있다. 그리고 [Yahoo!]등으로 검색된 검색엔진을 알려준다.

MetaCrawler(http://www.go2net.com/search.html)는 1995년부터 시작됐으며 데이터 베이스는 보유하고 있지 않다. 특징으로서, 데이터 베이스를 보유하고 있지 않지만 여러 검색엔진을 이용하여 광범위한 검색이 가능하다. 각 검색엔진 고유의 검색방법을 사용할 수 없으나, 희귀한 단어 등의 검색에 사용하면 좋다. MetaCrawler의 통계 관련 사이트 검색으로서, 검색어 statistics, mathematics로 검색시 각 대학의 홈페이지와 통계 관련 사이트를 검색할 수 있으며, AltaVista, Excite 등으로 검색된 검색엔진을 표시해준다.

Starting Point(http://www.stpt.com)는 170개 이상의 유명하고 고품질의 검색결과를 제공하는 검색도구를 사용하는 강력한 메타검색이다. Starting Point가 의뢰하는 검색엔진으로는 Yahoo, DejaNews, WebCrawler, Inktomi, Open Text, Savvy Search 등이 있다. 특징으로서, 170여개 이상의 좋은 검색결과를 제공하는 검색도구를 사용하는 검색엔진이다. Starting Point의 통계 관련 사이트 검색으로서, 검색어 statistics, mathematics로 검색시 각 대학의 홈페이지와 통계 관련 사이트를 검색할 수 있으며, 찾은 검색엔진을 표시해준다.

참고 문헌

1. 길명수, 반종오, 이황규, **인터넷 활용**, 연학사, 1998
2. 이형일, **인터넷의 기초**, 학문사, 1998
3. Kenneth C. Laudon, Jane Price Laudon, *Information System and The Internet*, Fourth Edition, The Dryden Press, 1998
4. *World Wide Web Journal*, Fourth International World Wide Web Conference Proceeding, O'Reilly & Associates, Inc. 1995