

# 인터넷 탐색엔진에 관한 연구

김 영 보\*

## 목 차

1. 서론
2. 이론적 배경
3. 연구의 방법
4. 분류체계의 분류
5. 컴퓨터·인터넷 분야 분류체계의 모형
6. 결론

## 1. 서 론

인터넷의 발전과 함께 최종이용자들이 직접 정보검색을 수행할 기회가 많아지고 있다. 그러나 컴퓨터, 네트워크 및 정보이론 등에 깊은 지식을 가지지 못한 최종이용자들이 전세계적으로 1천 6백만개<sup>1)</sup> 이상의 호스트가 접속되어 있는 인터넷 상에서 원하는 정보를 탐색하는 것이 쉬운 일이 아니다. 이러한 문제를 해결하기 위한 인터넷 도구로는, 고퍼(Gopher)를 검색하기 위한 베로니카(Veronica), 에프티피(FTP)를 검색하기 위한 아키(Archie), 월드 와이드 웹(WWW : World Wide Web)을 검색하기 위한 탐색엔진(Search Engine) 등이 있는데, 월드

---

\* 성균관대학교 대학원  
접수일자 1997년 12월 29일

1) Internet Domain Survey, 1997. <http://www.nw.com/zone/WWW/report.html>

와이드 웹의 성장과 함께 그 검색도구인 탐색엔진은 중요한 매개체로 자리잡고 있다.

탐색엔진은 크게 주제어 부여에 의한 검색을 제공하는 형태와 인터넷상의 서버들을 특정한 형태로 분류하여 계층구조로 탐색할 수 있도록 제공하는 형태로 나뉘어지는데, 특히 후자의 경우 체계적인 분류 이론의 도입이 미흡하여 실효성이 의문시되고 있다. 대부분의 탐색엔진은 불리언 논리식(Boolean logics)과 용어절단탐색, 개념에 의한 색인 등을 통해 검색효율을 높이고 있지만, 분류체계의 경우 CC, DDC, KDC, LCC와 같은 분류이론을 도입하지 않은 채 자체적인 임의의 체계를 따르는 경우가 대부분이어서 최종이용자들은 많은 탐색시간을 소요함에도 불구하고 원하는 정보를 찾지 못하는 경우가 많다. 특히, 컴퓨터·인터넷 분야의 경우 인터넷 탐색엔진의 대구분 항목 중 가장 많은 항목 수를 포함하고 있어 이에 대한 연구의 필요성이 크게 요구되고 있다.

탐색엔진에 대한 지금까지의 연구는 두 가지 방향에서의 접근을 시도하고 있다. 첫 번째는 주제어 부여에 의한 검색의 효율성을 높이기 위해 탐색엔진에 보다 많은 탐색 기법을 도입하고자 하는 시도이고, 두 번째는 탐색엔진상의 분류체계에 이론적 배경을 도입하고자 하는 시도이다. 이 중 이론적 배경의 도입에 의한 분류테이블의 구축에 대한 연구는 상대적으로 적은 편인데, 인터넷상에 흩어져 있는 수많은 자원을 체계적으로 분류하여 접근하도록 하는 것은 이용자의 탐색노력을 줄이고, 정보의 상·하위 관계를 파악하고 관련 주제를 함께 찾을 수 있는데 보다 유용한 방법이라고 여겨진다.

본 연구에서는 탐색엔진에 적용될 수 있는 컴퓨터·인터넷 분야 분류체계를 개발하는데 중점을 두어, 탐색엔진의 분류체계에 이론적 배경을 도입하고 각 탐색엔진의 분류체계를 비교·분석하여, 적절한 분류체계의 모형을 구축하고자 한다. 이런 목적을 위하여 본 연구에서 구체적으로 다루고자 하는 것은 다음과 같다.

첫째, 문헌 분류 이론과 인터넷 탐색엔진의 특성 분석을 통해 인터넷 탐색엔진에 적용될 수 있는 분류이론을 정립한다.

둘째, 기존의 인터넷 탐색엔진들이 제공하는 컴퓨터·인터넷 분야 분류체계를 정립된 분류이론에 근거하여 분석한다.

셋째, 컴퓨터·인터넷 분야에 적합한 인터넷 탐색엔진 분류체계의

모형을 구축한다.

## 2. 이론적 배경

### 2. 1 인터넷의 성장과 탐색엔진

인터넷 호스트와 도메인 수는 급격히 증가하고 있다. 국내의 경우, 1993년 12월과 1997년 4월 사이에 각각 1,135%, 8,851% 증가하였고(표 1 참조), 외국의 경우, 1993년 1월과 1997년 1월 사이에 각각 1,229%,

〈표 1〉 국내 인터넷 호스트, 도메인 수<sup>2)</sup>

연 도 별	호스트 수	도메인 수
1993년 12월	7,650	47
1994년 12월	14,681	192
1995년 12월	38,644	563
1996년 7월	58,350	1,377
1996년 12월	73,189	2,664
1997년 4월	86,805	4,160

〈표 2〉 해외 인터넷 호스트, 도메인수<sup>3)</sup>

연 도 별	호스트 수	도메인 수
1993년 1월	1,313,000	21,000
1993년 7월	1,776,000	26,000
1994년 1월	2,217,000	30,000
1994년 7월	3,212,000	46,000
1995년 1월	4,852,000	71,000
1995년 7월	6,642,000	120,000
1996년 1월	9,472,000	240,000
1996년 7월	12,881,000	488,000
1997년 1월	16,146,000	828,000

2) 한국인터넷정보센터, 1997, 국내 인터넷 호스트, 도메인 수 <http://www.krnic.net>

3) Internet Domain Survey, <http://www.nw.com/zone/WWW/report.html>

3.942% 증가하였다(표 2 참조).

이와 함께 호스트에 연결된 서버 역시 기하급수적으로 증가하고 있는데, 이러한 양적인 증가와 함께 최종이용자들의 정보검색을 더욱 어렵게 만드는 것은 역설적으로 월드 와이드 웹(WWW)의 보편화 추세이다. 월드 와이드 웹은 HTML(Hypertext Markup Language)을 기반으로 하이퍼텍스트 개념의 도입과 그림, 동영상, 소리 등의 멀티미디어 표현이 가능하다는 장점을 배경으로 인해 인터넷 표준으로 자리잡아가고 있는데, 그 성격상 이전의 고퍼, 에프티피와는 달리 각 서버들의 정보가 비선형 구조로 분산되어 존재하게 된다. 최종이용자들의 정보탐색을 돕기 위한 도구의 하나인 탐색엔진은 형태에 의해 ① 분류체계 제공형 엔진, ② 주제어 검색 제공형 엔진, ③ 분류체계와 주제어 검색 제공의 통합형 엔진, ④ 메타정보 엔진으로 구분할 수 있다. 분류체계 제공형 엔진은 인터넷상의 서버들의 정보를 수집한 뒤 계층 구조로 제공하는 형태의 엔진이고, 주제어 검색 제공형 엔진은 이용자가 부여한 주제어를 중심으로 데이터베이스 내의 해당자료를 탐색하는 형태의 엔진이다. 또한 통합형 엔진은 두 가지 서비스를 동시에 제공하는 형태를 말하며, 메타정보 엔진은 탐색용어를 일단 입력하고 여러 개의 탐색엔진에 동시에 적용하여 질의하는 방식을 취하는 형태를 말한다. 오늘날의 추세는 대부분의 탐색엔진들이 통합형 엔진으로 발전하고 있으며, 메타정보 엔진의 경우는 인터넷 서비스 제공자들의 사이트에서 제공되는 것이 보통이다. 탐색엔진은 월드 와이드 웹에서 빈번한 접속회수를 보이고 있는데, 외국의 경우 20여 개가, 국내의 경우 12개가 존재한다. 탐색엔진의 이용도가 높아지는 것은 그만큼 인터넷상에서 정보를 찾는 과정에 체계적인 도움이 필요함을 역설하는 것이라 하겠다.

## 2. 2 선행연구의 개관

정영미와 김성은은 탐색엔진의 성능을 평가하였다.<sup>4)</sup> 저자는 알타비스타(Alta Vista), 익사이트(Excite), 핫봇(HotBot), 인포식(Infoseek), 라이코스(Lycos), 마젤란(Magellan), 오픈 텍스트 인덱스(Open Text Index), 웹크롤러(WebCrawler), 야후(Yahoo!)를 분석

4) 정영미, 김성은, 1997. "WWW 탐색도구의 색인 및 탐색 기능 평가에 관한 연구." 한국문헌정보학회지 31:1, 153-184.

대상으로 선정하여, 각 탐색엔진의 색인 및 탐색 기능과 검색된 문서의 순위부여 방법을 비교한 후, 탐색실험을 통해 검색효율, 중복탐색의 정확도, 탐색결과의 유사도 등을 측정하였다. 연구의 결과에서 대부분의 탐색엔진이 질문의 성격과 작성된 탐색문에 따라 탐색결과에 있어 차이가 있는 것으로 나타났고, 재현율이 전반적으로 낮은 수치를 기록하였다. 또한 각 탐색도구간 유사도가 매우 낮았으며 탐색질문에 따라서도 탐색도구간의 분포가 다른 것으로 나타났다.

이명희는 주제별 디렉토리 및 키워드 검색엔진의 검색효율에 관한 연구를 수행하였다.<sup>5)</sup> 저자는 분류체계 제공형 엔진인 Yahoo!와 주제어 검색 제공형 엔진인 Alta Vista가 대학도서관 이용자들이 의해 제기된 탐색질문에 대해 얼마나 적합한 문헌을 탐색해 내는지 알아보기 위하여 탐색적 연구를 시도하였다. 저자에 의한 탐색결과는 검색된 문헌의 양, 검색된 적합문헌의 양, 재현율, 정확률의 측정기준에 의해 평가되었는데, Alta Vista는 특정적이고 기술적인 용어의 탐색에 적합한 반면 Yahoo!는 평이하고 일반적인 용어의 탐색에 적합한 것으로 드러났다.

스베노니우스(E. Svenonius)는 온라인 검색에 있어서의 분류체계의 이용에 관해 연구하였다.<sup>6)</sup> 저자는 분류체계를 이용하는 것이 적합율과 재현율을 높이고 이용자의 시간을 절약하는데 유용하며, 특히 통계자료와 같은 비 서지데이터베이스의 설계와 관련 색인, 유사 초록, 자연어 용어의 배치 시에 중요한 방법이라고 보았다. 이는 분류체계가 용어간 상관관계를 계층적으로 보여주며 의미론적 브라우징에 도움이 되기 때문이며 따라서, 분류체계의 이론과 실제 적용이 필요하다고 보았다.

비자인-고츠(D. Vizine-Goetz)는 1983년 스베노니우스의 연구를 이어받아 인터넷 자원에 대한 도서관 분류표의 적용을 연구하였다.<sup>7)</sup>

저자는 DDC와 LC를 선정하여 탐색엔진 Yahoo!의 분류체계중 1-10, 35-45 범주와 비교·분석하였는데, 각 항목들의 용어와 하위 구성에 포함된 자원의 수를 조사하여 항목간 균형성과 적절성을 밝혀냈다.

5) 이명희, 1997. "네트워크 데이터베이스에서의 주제별 디렉토리 및 키워드 검색엔진의 검색효율에 관한 탐색적 연구." 한국문헌정보학회지 31:2, 177-197.

6) E. Svenonius, 1983. "Use of classification in online retrieval." *Library Resources & Technical Services*, 27:1, 76-80.

7) D. Vizine-Goetz, 1996. "Using library classification schemes for Internet resources (Position Paper)." *Proceedings of the OCLC Internet Cataloging Colloquium*, San Antonio, Texas, January 19, 1996. Dublin, Ohio: OCLC. <http://www.oclc.org/oclc/man/colloq/v-g.htm>

인포마인(INFORMINE) 프로젝트는 미의회도서관 주제명 표목표(Library of Congress Subject Headings : LCSH)를 이용하여 인터넷 자원의 분류체계를 구축하였다.<sup>8)</sup> 대주제로 ① Biological, Agricultural & Medical Sciences ② Government Information ③ Introductory Resources : K-12 ④ Introductory Resources : University ⑤ Internet Enabling Tools ⑥ Maps & GIS ⑦ Physical Sciences, Engineering, Computing & Math ⑧ Regional & General Interest ⑨ Social Sciences & Humanities ⑩ Visual & Performing Arts을 채택했고, 각각의 아래에 하위 주제를 자모순으로 배열하여 접근하도록 하였다. 이를 통해 상위 주제와 특정 주제에 접근이 가능한 메타-주제 용어로도 접근이 가능하도록 하였다.

달버그(I. Dahlberg)는 네트워크 환경에서의 분류이론 적용에 관한 연구를 수행했다.<sup>9)</sup> 저자는 LC와 DDC를 선정하여 새로운 분류체계로서 적합한지를 분석하였는데, 분류이론의 적용은 각 주제의 분석과 추적 및 계층의 구분에 유용하며 또한 기존에 분류되어 있는 자료의 네트워크 상 재조직에도 필요하다고 보았다.

월리스(J. Wallis)와 버든(P. Burden)은 월드 와이드 웹 자원에 대한 분류체계 기반 탐색을 분석하였다.<sup>10)</sup> DDC를 응용한 WWlib과 같은 분류체계기반 탐색엔진이 특히 텍스트기반 탐색에서 우수한 성능을 보였는데, 이를 통해 Automated Classification Engine(ACE) 시스템을 제안하였다. 저자들은 분류체계 기반 탐색엔진 설계가 많은 탐색엔진에 적용되어야 한다고 보았다.

핏코우(J. E. Pitkow)와 케호(C. M. Kehoe)는 월드 와이드 웹 이용 자들에 대한 조사연구를 수행하였다.<sup>11)</sup> 이용자들을 연령별, 성별, 수입, 지역, 인종, 직업별로 구분하여 조사하였고, 이들이 접속에 사용하는 비용과 시간, 사용하는 시스템과 웹 브라우저, 탐색하는 자원의 종류,

8) INFORMINE, 1997. <http://lib-www.ucr.edu>

9) I. Dahlberg, 1995. "The Future of Classification in Libraries and Networks, a Theoretical Point of View," *Cataloging & Classification Quarterly*, 21:2, 23-35.

10) J. Wallace and P. Burden, 1995. Toward a classification-based approach to resource discovery on the Web. Wolverhampton: University of Wolverhampton, School of Computing and Information Technology. <http://www.scit.wlv.ac.uk/wwlib/position.html>

11) J. E. Pitkow and C. M. Kehoe, 1997. Result from the Third WWW User Survey. <http://www.w3.org/pub/www/Journal/1/pitkow.107/paper/107.html>

HTML에 대한 학습정도가 분석되었다. 연구결과는 이용자들이 “질문-응답-확인/재질문”의 형태로 인터넷 자원을 탐색하는 것으로 나타났다.

### 2. 3 선행연구의 분석

인터넷 탐색엔진의 분류체계와 관련된 선행연구들을 개관한 결과, 지금까지의 연구는 탐색엔진의 검색 효율성을 측정하려는 시도, 온라인 목록의 분류체계 구축의 효율성에 관한 분석, DDC, LC, UDC등 분류이론의 적용을 통해 탐색엔진과는 별도의 시스템을 구축하고자 하는 시도가 주축을 이루고 있다. 탐색엔진의 분류체계에 다양한 분류이론을 적용하여 발전시키려는 연구는 부족하며, 더욱이 컴퓨터·인터넷 분야에 대한 연구는 매우 부족한 편이다.

이에 본 연구는 바로 이 점에 문제의식을 갖고, 탐색엔진에 적용될 수 있는 컴퓨터·인터넷 분야 분류체계를 개발하고자 한다. 구체적으로는 탐색엔진의 분류체계에 이론적 배경을 도입하고 각 탐색엔진의 분류체계를 비교·분석하여, 적절한 분류체계의 모형을 구축하고자 한다.

## 3. 연구의 방법

### 3. 1 연구 대상 및 표본의 선정

본 연구에서는 분석의 대상으로 분류체계 제공형 탐색엔진과 통합형 탐색엔진 중 분류체계 제공에 중점을 두고 있는 국내외의 대표적인 엔진을 선정하여 분석한다. 분석의 대상은 컴퓨터·인터넷 분야의 항목수와 탐색엔진의 이용도를 고려하여 아래와 같이 선정하였다. 국내의 경우는 최초로 탐색엔진 서비스를 제공하기 시작하여 높은 이용도를 보이고 있는 심마니와 분류체계에 보다 역점을 두고 있는 정보탐정을 선정하였고, 해외의 경우는 통합형 엔진 중 가장 높은 이용도를 보이는 Excite와 Yahoo!를 선정하였는데, Yahoo!의 경우는 1997년 10월부터 제공되고 있는 한국어 사이트 Yahoo Korea!로 대치하였다. Yahoo!와 Yahoo Korea!는 동일한 항목을 가지고 있다.

. 심마니 (<http://simmany.hnc.net>)

- . 정보탐정 (<http://idetect.kotel.co.kr>)
- . Excite (<http://www.excite.com>)
- . Yahoo Korea! (<http://www.yahoo.co.kr>)

### 3. 2 연구의 방법

본 연구에서는 연구방법으로서 비교 분석과 이론적 모형 구축을 위한 문헌연구를 채택한다.

### 3. 3 연구의 제한점

본 연구에서는 문헌연구를 통해 인터넷 탐색엔진의 컴퓨터·인터넷 분야 분류체계의 모형을 구축하므로, 구축된 모형의 실효성 분석을 위해서 본 연구의 결과를 이용한 탐색엔진의 구축과 이용자 연구가 필요하다.

## 4. 분류체계의 분석

### 4. 1 문헌 분류이론과 탐색엔진의 특성 분석

마르셀로(R. Marcello)와 뉴튼(R. Newton)은 분류의 가장 중요한 목적은 정보 탐색자에게 최대한 유용한 주제의 배열을 작성하고 제공하기 위한 것으로 정의하고 있다.<sup>12)</sup> 다시 말해 수집된 자료를 이용자에게 효과적으로 제공하는데 분류의 목적이 있다 하겠다. 따라서 그 기준은 이용자가 가장 쉽게 자료에 접근하는 방법에 중점을 두어야 하며, 이를 위해서는 분류체계 구성상의 기본조건이 필요하다.

이러한 기본조건으로서 김명옥은 시스템 중심적인 기본조건을 제기하였다.<sup>13)</sup> 그에 의하면, 분류이론은 첫째, 보편성, 객관성을 갖춘 학문의 분류에 순응할 것, 둘째, 자료에 표현되는 과거, 현재, 미래의 모든 주제를 처리할 수 있는 포괄성, 정밀성, 가능성을 가질 것, 셋째, 구분원리가

12) R. Marcello & R. Newton, 1994. A New Manual of Classification. Gower Publishing, p. 3.

13) 김명옥, 1989. 자료분류법. 구미무역(주)출판부, p. 34.



일관적이고, 체계적일 것, 넷째, 분류표가 구체적이고 정밀할 것, 다섯째, 분류명사가 명확할 것 등을 들고 있다. 이러한 조건들은 분류체계 구축에 있어 핵심적인 것이라 하겠다.

김명옥이 제기한 시스템적 측면의 분류이론은 분류체계 구축에 핵심적이거나, 이용자 중심적 측면에서는 다소 부족하다 하겠다. 이용자중심적 측면에서의 기본조건으로 생각해 볼 수 있는 것은, 첫째, 하이퍼텍스트의 응용을 통한 탐색시 공간적 거리 단축, 둘째, 사용빈도가 높은 주제어의 우선 배치를 통한 탐색노력의 단축, 셋째, 전문용어와 비전문용어의 혼용 등을 들 수 있다. 하이퍼텍스트의 응용은 주제의 지나친 세분에 의한 탐색거리와 분류체계 구축과 색인시의 노력을 줄일 수 있고, 사용빈도가 높은 주제어의 우선 배치는 시스템적 관점에서 부적절하나, 탐색자의 탐색노력을 줄이는데 효과적이다. 또한 전문용어와 비전문용어의 혼용을 통해 보다 친숙한 이용자 인터페이스를 구축할 수 있을 것이다.

이상의 분류체계 구축시 기본조건을 정리하면 다음과 같다.

- ① 포괄성 : 분류체계가 컴퓨터·인터넷 분야의 주제를 포괄하여 망라하며, 이후의 새로운 분야에 대해 능동적으로 대처 가능한가?
- ② 논리성 : 분류체계가 컴퓨터·인터넷 분야의 주제의 상·하위간 인과관계, 항목간의 균형성과 일관성에서 논리적이고 객관적인가?
- ③ 명확성 : 분류체계의 주제어가 컴퓨터·인터넷 분야의 주제를 명확하고 정밀하게 표현하며, 탐색자의 인지 가능성은 높은가?
- ④ 효율성 : 분류체계가 컴퓨터·인터넷 분야의 주제를 색인하는데 적절하며, 탐색시의 노력을 줄이는데 효율적인가?

#### 4. 2 분류체계의 포괄성 비교 분석

<표 3>은 대구분 항목 수가 가장 적은 심마니를 기준으로 정보탐정, Excite, Yahoo Korea!의 대응되는 항목을 매칭시킨 것이다. 이를 통해 분류체계의 포괄성을 비교 분석하였는데, 대구분 항목의 수는 Yahoo Korea!(33개), 심마니(17개), 정보탐정(11개), Excite(9개) 순으로 나타났고, 각각의 하위구분 항목의 수는 Excite(167개: 중복 항목 제외), Yahoo Korea!(126개: 중복 항목 제외), 심마니(37개), 정보탐정(30개)의 순으로 나타났다.

#### 4. 2. 1 심마니의 포괄성 분석

심마니의 경우, 대구분 항목의 수는 17개였으나 하위구분 항목의 수가 37개여서, 대구분 항목이 가장 적은 Excite의 하위구분 항목에 비해 월등히 적었다. 이는 컴퓨터·인터넷 분야의 주제어를 많이 포괄하고 있지 않음을 나타낸다. 분류의 전개는 두 단계의 계층구조였고, 대구분 항목 중 절반이 하위구분 항목을 가지고 있지 않아 새로운 항목에 대한 전개 가능성이 높지 않았다.

#### 4. 2. 2 정보탐정의 포괄성 분석

정보탐정의 경우, 대구분 항목의 수는 11개로 Excite에 비해 많았으나, 하위구분 항목의 수가 30개로 가장 적어 심마니와 마찬가지로 컴퓨터·인터넷 분야의 주제어를 많이 포괄하지 않고 있는 것으로 나타났다. 분류의 전개는 두 단계의 계층구조였고, 11개의 대구분 항목 중 3개가 하위구분 항목을 가지고 있지 않아 새로운 항목에 대한 전개 가능성이 높지 않았다.

#### 4. 2. 3 Excite의 포괄성 분석

Excite의 경우, 대구분 항목의 수는 탐색엔진 중 가장 적었으나, 하위구분 항목의 수는 167개로 가장 많은 주제어를 포괄하고 있었다. 분류의 전개는 다섯 단계의 계층구조였고, 대구분 항목 중 하위구분 항목을 가지지 않은 항목이 없고, 치밀하고 방대하게 분류하고 있어 새로운 항목에 대한 전개 가능성이 높았다.

#### 4. 2. 4 Yahoo Korea!의 포괄성 분석

Yahoo Korea!의 경우, 대구분 항목의 수는 33개로 가장 많았으나, 하위구분 항목의 수는 126개로 Excite에 비해 오히려 적었다. 분류의 전개는 다섯 단계의 계층구조였고, 33개의 대구분 항목 중 하위구분 항목을 가지지 않은 항목은 11개 였으나, 이들은 이용빈도가 높은 주제어를 세도우로 처리한 것이어서 하위구분 항목을 가지지 않았다고 보기 어렵다. 따라서, Yahoo Korea!의 새로운 항목에 대한 전개 가능성은 높았다.

#### 4. 3 분류체계의 논리성 비교 분석

〈표 3〉 분류체계의 대구분 항목

심마니	정보탐정	Excite	Yahoo Korea!
BBS	BBS, 채팅	Internet	월드와이드웹 (WWW)
검색시스템 (검색엔진)	인터넷 검색		링크모음
서비스(전문업체)			인터넷
인터넷 교실	인터넷 강좌		통신, 네트워크
자료실			참고자료와 정보
카페	인터넷 서비스		
기타	전산학, 컴퓨터공학	Computer Science	2000년 표기문제
			공학
			반도체
			보안과 암호화
			전자통신
		Networking	하드웨어
			퍼스널 컴퓨터
			표준
			한글처리
			휴대형 컴퓨터
단체, 기관	컴퓨터/인터넷 기업, 기관	News & Magazine	기관, 단체
회사(업체)			동호회
			취업
신문	컴퓨터/인터넷 간행물	회사	
웹매거진		유머	
출판류	컴퓨터/인터넷 뉴스, 행사	잡지	
행사, 대회		전자출판	
매킨토시(MAC)	운영체제		컨벤션과 회의
프로그래밍	소프트웨어	Shareware	콘테스트
		Software	운영체제
		Programming	게임
		Cyber Culture	
CAD		Graphics	프로그래밍 언어
		Multimedia	사이버문화
			그래픽
			멀티미디어
			예술
			음악

4. 3 .1 심마니의 논리성 분석

심마니의 분류는 특정한 원칙을 따르고 있지 않았다. 심마니는 상 하 위간 인과관계와 일관성 및 항목간 균형성에서 적절하지 못한 부분을 보여주었다. 대구분의 CAD 항목, 매킨토시 항목, 신문 항목, 웹매거진 항목, 카페 항목, 보안 항목 등은 다른 탐색엔진의 하위구분에 포함된

항목들을 대구분으로 설정한 경우였는데, (표 3 참조) Yahoo Korea!와 달리 이용빈도에 따라 대구분에 설정한 것으로 보기 어렵다. 하위구분 항목 중 자료실 항목의 하위인 정보서비스 항목과 추천사이트 항목은 검색 시스템 항목의 하위구분에 포함되는 것이 적절하다. 또한 컴퓨터·인터넷 분야의 학문분류와 달리 컴퓨터공학, 네트워크, 프로그래밍 등의 항목은 존재하지 않았는데, 이들 항목의 신설이 필요하다. (표 4 참조)

#### 4. 3. 2 정보탐정의 논리성 분석

정보탐정의 분류는 컴퓨터·인터넷 분야의 주제에 의한 전개와 정보

〈표 4〉 심마니의 논리성 분석

현재의 항목	개선안
없음	컴퓨터공학 항목 신설
없음	네트워크 항목 신설
없음	소프트웨어 항목 신설
없음	프로그래밍 항목 신설
CAD	그래픽 항목의 신설, 그 하위에 구분
매킨토시	컴퓨터-소프트웨어 항목의 신설, 그 하위에 구분
신문	출판류의 하위에 구분
웹매거진	출판류의 하위에 구분
카페	서비스(전문업체)의 하위에 구분
보안	서비스(전문업체)의 하위에 구분
정보서비스	검색시스템(검색엔진)의 하위에 구분
추천사이트	검색시스템(검색엔진)의 하위에 구분

원에 의한 전개를 병행하고 있었다. 정보탐정은 상·하위간 인과관계와 일관성 및 항목간 균형성에서 적절하지 못한 부분을 보여주었다. BBS, 채팅 항목은 컴퓨터/인터넷 기업, 기관 항목의 하위구분에 포함되고, 운영체제 항목은 전산학, 컴퓨터공학 항목의 하위에 포함되는 것이 적절하며, 전산학, 컴퓨터 공학 항목은 하위 항목으로 많은 항목을 포함할 수 있는 항목임에도 불구하고, 하위 항목을 가지고 있지 않았다. 컴퓨터와 인터넷 부분이 통합되어 기업, 기관 항목과 간행물 항목, 뉴스,

행사 항목을 나타내고 있는데, 컴퓨터/인터넷 부분은 각기 다수의 항목을 포함하며 독립되어야 할 항목이다. 운영체제 항목의 하위에 네 가지의 주요 운영체제를 명시하고 있으나, NextStep, Be OS 등의 기타 운영체제에 대한 항목 부여가 되어 있지 않다. 인터넷 서비스 항목의 인터넷길잡이사이트와 서버가이드는 인터넷 강좌의 하위구분으로 포함되어야 적절하다. 소프트웨어 항목의 하위 항목은 게임과 Shareware & Freeware로 구분되어 있는데, 이는 소프트웨어의 성격으로 구분하는 방법이나 상용 프로그램/세어웨어/공개 프로그램 등으로 구분하는 방법이 적절하다. 컴퓨터/인터넷 간행물 항목의 하위구분은 단행본, 정기간행물 형식으로 구분하거나 학술지, 잡지, 단행본 형식으로 구분하는 것이 적절하다. (표 5 참조)

4. 3. 3 Excite의 논리성 분석

Excite의 분류는 컴퓨터·인터넷 분야의 주제에 의한 전개를 따르고

<표 5> 정보탐정의 논리성 분석

현재의 항목	개선안
BBS, 채팅	컴퓨터/인터넷 기업, 기관의 하위에 구분
운영체제	전산학, 컴퓨터공학의 하위에 구분
전산학, 컴퓨터공학	하위 세구분
컴퓨터, 인터넷 관련항목	컴퓨터와 인터넷으로 각기 독립
운영체제	하위에 기타 항목 신설
인터넷길잡이사이트	인터넷 강좌의 하위에 구분
소프트웨어	하위 항목 재 분류
컴퓨터/인터넷 간행물	하위 항목 재 분류
서버가이드	인터넷 강좌의 하위에 구분

있다. Excite는 상·하위간 인과관계, 항목간 균형성 및 일관성에서 우수해서 심마니와 정보탐정에 비해 수정되어야 할 항목의 수가 적었다. 다만, Graphics 항목의 하위에 존재하는 Archives와 Clip Art는 통합되어야 하고, Hardware 항목의 하위구분인 Personal Computers 항목은 Companies 하위로 포함되어야 한다. Shareware 항목이 독립되어 있는 것은 부적절한데, 이는 Software 항목의 하위에 포함되어야 하고, Freeware 항목이 추가되어야 한다.

〈표 6〉 심마니의 논리성 분석

현재의 항목	개선안
Archives	Clip Art와 통합
Personal Computers	Companies의 하위 항목에 구분
Shareware	Software의 하위 항목에 구분
없음	Software의 하위 항목에 Freeware 신설

#### 4. 3. 4 Yahoo Korea!의 논리성 분석

분류의 전개는 컴퓨터·인터넷 분야의 주제에 의한 전개와 정보원에 의한 전개를 병행하고 있다. Yahoo Korea!는 상·하위간 인과관계 및 항목간 균형성 및 일관성에서 비교적 우수했으나, 적절하지 않은 부분을 보여주었는데, 잡지 항목은 독립시키지 않고 그 성격에 따라 인터넷 항목과 전자출판 항목, 컨벤션과 회의 항목, 컴퓨터 공학 항목 등으로 세분되어 편입되어야 한다. 휴대용 컴퓨터 항목은 하드웨어 항목의 하위에 포함되어야 한다. (표 7 참조) 특기할 점은, 주제에 의한 전개와 정보원에 의한 전개를 주로 하고, 여기에 탐색빈도가 높은 주제어를 OS/2의 세도우 개념으로 처리하여 병행하는 형식으로 작성되었다는 점인데, 이는 항목간 균형성과 일관성에서 적절하지 않다.

〈표 7〉 Yahoo Korea!의 논리성 분석

현재의 항목	개선안
잡지	인터넷, 전자출판, 컨벤션과 회의, 컴퓨터 공학의 하위에 세분하여 구분
휴대용 컴퓨터	하드웨어의 하위구분

#### 4. 4 분류체계의 명확성 비교 분석

##### 4. 4. 1 심마니의 명확성 분석

심마니가 선정한 주제어는 컴퓨터·인터넷 분야 주제를 명확하게 표

현하지 못하고 있어, 탐색자의 인지가능성이 낮다. 대구분 항목의 '서비스(전문업체)'는 '인터넷 서비스 전문업체'로 변경되어야 하며, '인터넷 교실' '카페'는 각각 '참고정보' '인터넷 카페'로 변경되어야 한다. 서비스(전문업체) 항목 하위구분의 '전자우편'은 'E-mail'로 변경되어야 하는데, 인터넷의 'E-mail'은 컴퓨터 통신의 '전자우편'과 구분되어지기 때문이다. 출판류 항목 하위의 'SW'와 회사(업체) 항목 하위의 '유통'은 무엇을 의미하는 지 정확하지 않다. (표 8 참조)

〈표 8〉 심마니의 명확성 분석

현재의 항목	개선안
서비스(전문업체)	인터넷 서비스전문업체
인터넷 교실	참고정보
카페	인터넷 카페
전자우편	E-mail
SW	워드프로세서/DTP 프로그램/Autoring 툴

#### 4. 4. 2 정보탐정의 명확성 분석

정보탐정이 선정한 주제어는 컴퓨터·인터넷 분야의 주제를 명확하게 표현하고 있어 탐색자의 인지가능성이 높다. 다만, 대구분 항목의 '전산학, 컴퓨터공학'은 '컴퓨터공학'으로 바뀌어야 하는데, 전산학은 컴퓨터공학과 구분되는 분야이다. '인터넷 서비스'는 '인터넷 서비스 전문업체'로 바뀌어야 한다. (표 9 참조)

〈표 9〉 정보탐정의 명확성 분석

현재의 항목	개선안
전산학, 컴퓨터공학	컴퓨터공학
인터넷 서비스	인터넷 서비스 전문업체

#### 4. 4. 3 Excite의 명확성 분석

Excite의 주제어는 컴퓨터·인터넷 분야의 주제를 명확하게 표현하고 있으며, 탐색자의 인지 가능성이 높다. 전체적으로 컴퓨터·인터넷 분

야의 주제어를 치밀하고 방대하게 전개하고 있는데, Computer Science 항목의 하위구분인 Secure HTTP 항목, Graphics 항목의 하위구분인 Validation/Checkers 항목, Hardware 항목의 하위구분인 Semiconductors 항목 등 다른 탐색엔진에서 사용하지 않는 전문용어를 사용하여 정밀성을 기하고 있다.

#### 4. 4. 4 Yahoo Korea!의 명확성 분석

Yahoo Korea!의 주제어는 컴퓨터·인터넷 분야의 주제를 명확하게 표현하여 탐색자의 인지 가능성이 높다. 전문적인 용어와 속어를 병행해 사용하고 있으며, 주제어의 수도 풍부하다. 다만, 미국의 Yahoo! 서비스의 주제어를 우리말로 옮기는 과정에서 적절하지 못하게 번역된 주제어들이 존재한다. 멀티미디어 항목의 하위구분인 '동영상 GIF'는 '움직이는 GIF'로, 소프트웨어 항목의 하위구분인 '엔지니어링'은 '공학'으로, 잡지 항목의 하위구분인 '웹진'은 '웹매거진'으로 번역되어야 한다.

〈표 10〉 Yahoo Korea!의 명확성 분석

현재의 항목	개선안
동영상 GIF	움직이는 GIF
엔지니어링	공학
웹진	웹매거진

#### 4. 5 분류체계의 효율성 분석

##### 4. 5. 1 심마니의 효율성 분석

심마니의 분류체계는 컴퓨터·인터넷 분야 주제를 색인 하는데 적절하지 않으며, 탐색자의 노력을 줄이는데 적절하지 않다. 심마니는 상하위간 인과관계와 일관성에서 적절하지 않아 색인에 효율적이지 못하고, 선정한 주제어가 탐색빈도가 낮거나 중요하지 않다. 또한, 항목의 수가 적어 단일 항목에서 다수의 인터넷 자원이 검색되어 탐색자의 노력을 줄이는데 효율적이지 못하다.



#### 4. 5. 2 정보탐정의 효율성 분석

정보탐정은 컴퓨터·인터넷 분야의 주제를 색인하는데 효율적이나, 탐색자의 노력을 줄이는데 효율적이지 못하다. 정보탐정은 상·하위간 인과관계와 일관성 및 주제어의 선정이 우수해서 색인에 효율적이나, 항목의 수가 적어 단일 항목에서 다수의 인터넷 자원이 검색되어 탐색자의 노력을 줄이는데 효율적이지 못하다.

#### 4. 5. 3 Excite의 효율성 분석

Excite는 컴퓨터·인터넷 분야의 주제를 색인하는데 효율적이나, 탐색자의 노력을 줄이는데 효율적이지 못하다. Excite는 상·하위간 인과관계와 일관성 및 주제어의 선정이 우수해서 색인에 효율적이나, 계층구조가 5단계로 구분되어 탐색시간이 많이 걸리고, 기계적인 계층구조로 분류되어 탐색거리가 길어 탐색자의 노력을 줄이는데 효율적이지 못하다.

#### 4. 5. 4 Yahoo Korea!의 효율성 분석

Yahoo Korea!는 컴퓨터·인터넷 분야의 주제를 색인하고 탐색자의 노력을 줄이는데 효율적이다. Yahoo Korea!의 특징은 탐색빈도가 높은 주제어를 대구분 항목에 포함시키는 것과 웹브라우저의 북마크와 유사 개념을 도입했다는 것, 그리고 OS/2의 세도우와 유사한 개념을 도입했다는 데 있다. 북마크 개념은 동일한 항목간을 한번에 이동할 수 있도록 한 것인데, 다른 위치의 동일 항목으로 이동하므로써 전후관계를 확장할 수 있고, 탐색 거리를 단축할 수 있다는 장점이 있다. 세도우 개념은 하나의 항목을 다른 위치로 이동할 때, 혹은 동일 항목이 여러 군데에 나올 때 하나의 항목만을 분류하고, 다른 항목들은 그곳으로 이동할 수 있는 '그림자'만 제공하는 것이다. 이 두 가지 개념이 적용된 항목은 주제어의 오른쪽에 @마크가 붙어있는 것들인데, 색인시 중복색인을 피하고, 탐색자가 주제어의 전후관계를 살펴거나 항목간을 '점프'할 때 효과적인 역할을 수행한다. Yahoo Korea!의 두 가지 특징은 하이퍼텍스트의 특징을 잘 응용한 예라 하겠다.

### 5. 컴퓨터·인터넷 분야 분류체계의 모형

탐색엔진들의 컴퓨터·인터넷 분야 분류체계를 비교 분석하여 분류체계의 모형을 구축하였다. 분류체계의 구축은 컴퓨터·인터넷 분야의 주제에 의한 전개와 정보원의 형태에 의한 전개를 주로 하였고, Yahoo Korea!와 마찬가지로 북마크 개념과 세도우 개념을 응용하였으며, 중복되는 항목은 제외하였다. 이는 전통적인 분류체계의 포괄성, 논리성, 명확성을 유지하면서, 효율성을 높이기 위해서이다. 작성된 분류체계는 기계적인 계층구조의 문제점을 보완하여, 탐색자로 하여금 자신이 원하는 정보를 주제별 혹은 정보원별로 탐색하면서, 동시에 항목간을 비선형으로 이동하여 탐색 거리를 줄이고, 주제의 상·하위, 동위 관계를 파악하기 쉽게 하여 탐색노력을 줄일 수 있다.

주제에 의한 전개를 따라 대구분 항목으로 선정한 것은 멀티미디어, 컴퓨터 공학, 인터넷인데, 그 이유는 컴퓨터 공학과 인터넷의 양쪽 영역을 포괄하는 멀티미디어 항목은 독립되어야 하기 때문이다.

정보원에 의한 전개를 따라 대구분 항목으로 선정한 것은 기관·단체, 통합자료실, 전문업체, 출판물인데, 심마니, 정보탐정, Yahoo Korea!의 경우를 참고하였다.

하위구분에 포함된 항목 중 사용빈도가 높은 게임 항목, 웹페이지 디자인 항목, E-mail 항목을 세도우 형태로 대구분에 위치시켰는데, 심마니, 정보탐정, Excite, Yahoo Korea!의 항목 출현 빈도를 참고하였다. 완성된 분류체계의 형태는 다음과 같다.

〈표 11〉 컴퓨터·인터넷 분야 분류체계의 모형

게임@			
기관·단체	대학	학과	
		연구소	
	동호회	멀티미디어	
		인터넷	
		컴퓨터공학	소프트웨어 하드웨어
		통신·네트워크	
	연구소	멀티미디어	
		인터넷	
		컴퓨터공학	소프트웨어 하드웨어
		통신·네트워크	
	협회	멀티미디어	
		인터넷	
컴퓨터공학			
통신·네트워크			
멀티미디어	가상현실	기관·단체	동호회 연구소
		전문업체	
	그래픽	기관·단체	동호회 연구소
		전문업체	
		출판물	단행본 잡지
	멀티미디어	기관·단체	동호회 연구소
		전문업체	
		출판물	단행본 잡지
	사운드	기관·단체	동호회 연구소
		전문업체	
인터넷	기타	FTP	Archie
		Gopher	Veronica
		IRC	
		Telnet	
		Usenet	
	오락	게임	
		방송	

인터넷	오락	우스개		
	인터넷 탐색	국내 탐색엔진		
		해외 탐색엔진		
	전문업체	교육		
		웹서비스	웹 서버 구축	
			웹 호스팅	
			접속 서비스	
	인터넷 광고			
	인터넷 카페			
	정책			
출판물	단행본			
	잡지			
	학술지			
E-mail@	Free E-mail			
	Mailing Lists			
웹페이지 디자인@	디자인 가이드			
	이미지			
	HTML/SGML			
전문업체	멀티미디어	그래픽		
		멀티미디어		
		사운드		
	인터넷	게임		
		교육		
		웹서비스	웹 서버 구축	
			웹 호스팅	
			접속 서비스	
	광고			
	인터넷 카페			
전자출판	소프트웨어 개발	워드프로세서		
		DTP 프로그램		
	글자체 개발			
컴퓨터공학	소프트웨어			
	통합솔루션			
	하드웨어			
출판물	단행본	게임		
		멀티미디어	그래픽	
			멀티미디어	
		사운드		
	인터넷			
컴퓨터공학	소프트웨어			
	하드웨어			

출판물	잡지	게임	
		멀티미디어	
		인터넷	
		컴퓨터 일반	
출판물	학술지	멀티미디어	그래픽
			멀티미디어 사운드
		인터넷	
		컴퓨터 공학	소프트웨어
통합솔루션 하드웨어			
컴퓨터공학	소프트웨어	게임	
		공학 관련	
		데이터베이스	
		멀티미디어	그래픽
			멀티미디어
			사운드
		운영체제	기타
			Linux
			Mac OS
			OS/2
			Unix
		유틸리티@	Windows 95/NT
			바이러스 시스템유틸리티
		인터넷@	
		전자출판	
		출판물	단행본
			잡지
학술지			
통신·네트워크			
프로그래밍	기타		
	C/C++		
	Java		
	Visual Basic		
	VRML		
통합솔루션	보안		
	시스템보안 전자화폐		
하드웨어	내부장치	기타	
		메인보드	

컴퓨터공학	하드웨어	내부장치	Memory	
		주변장치	디지털 카메라	
			랜 카드	
			모니터	
			모뎀	
			스캐너	
			입력장치	
			프린터	
			하드디스크	
			CD-ROM 드라이브	
			컴퓨터 완제품	노트북
		Desktop Computer		
		NC		
			Palmtop,PDA	
통합자료실 (세어웨어/프리웨어)	게임	기타		
		머드		
		롤플레이잉		
		시뮬레이션		
		아케이드		
	멀티미디어	그래픽	기타 운영체제	
			Linux	
			OS/2	
		멀티미디어	Windows 95/NT	
			기타 운영체제	
			Linux	
		사운드	OS/2	
			Windows 95/NT	
			기타 운영체제	
	인터넷@	기타	Linux	
			OS/2	
			Windows 95/NT	
FTP				
웹 브라우저		IRC		
		Telnet		
		Usenet		
		기타 운영체제		
		Linux		
			Windows 95/NT	
			Unix	

통합자료실 (세어웨어/ 프리웨어)	인터넷@	웹페이지 디자인	기타 운영체제
			Linux
			Windows 95/NT
			Unix
	컴퓨터공학	E-mail	기타 운영체제
			Windows 95/NT
		디바이스 드라이버	
		롬 바이오스	
		백신 프로그램	
		운영체제	
유틸리티@			
프로그래밍 개발툴			
웹 페이지 디자인@			
E-mail@			

## 6. 결 론

인터넷 탐색엔진의 분류체계에 적용 가능한 분류이론을 확립하고, 인터넷 탐색엔진의 컴퓨터·인터넷 분야 항목을 비교 분석하여 컴퓨터·인터넷 분야의 분류에 적합한 탐색엔진 분류체계의 모형을 구축해 보았다. 본 연구를 통해 밝혀진 주요 내용을 정리하면 다음과 같다.

(1) 분류의 가장 중요한 목적은 정보 탐색자에게 유용한 주제를 배열하고 제공하기 위한 것으로, 인터넷 탐색엔진의 분류체계가 갖추어야 할 분류기준은 주제의 포괄성, 분류체계의 논리성, 주제어의 명확성, 분류체계의 효율성 등이다.

(2) 인터넷 탐색엔진 분류체계의 포괄성은 각 탐색엔진의 대구분 항목과 하위구분 항목의 주제어 수를 통해 비교 분석되었다. 분석결과 Excite의 포괄성이 가장 우수한 것으로 나타났다.

(3) 인터넷 탐색엔진 분류체계의 논리성은 각 탐색엔진의 상·하위 간 인과관계, 항목간 균형성과 일관성에서 분석되었다. 분석결과 Excite의 논리성이 가장 우수한 것으로 나타났다.

(4) 인터넷 탐색엔진 분류체계의 명확성은 주제어의 명확성과 정밀성, 탐색자의 인지 가능성에 의해 분석되었다. 분석결과, 정보탐정과 Excite의 명확성이 가장 우수한 것으로 나타났다.

(5) 인터넷 탐색엔진 분류체계의 효율성은 주제 색인시의 적절성과 탐색 노력 감소의 효율성에 의해 분석되었다. 분석결과, Yahoo Korea!의 효율성이 가장 우수한 것으로 나타났다.

(6) 컴퓨터·인터넷 분야에 적합한 인터넷 탐색엔진의 분류체계 모형은 주제의 포괄성, 분류체계의 논리성, 주제어의 명확성, 분류체계의 효율성 등의 제고를 위해 구축되었다. 모형은 주제에 의한 전개와 정보원의 형태에 의한 전개를 주로 하였고, 북마크 개념과 세도우 개념을 응용하였다.



## 참 고 문 헌

- 강덕수, 1992. 이용자 인터페이스 향상을 위한 온라인 목록의 기능에 관한 연구. 이화여자대학교 석사학위논문. 1992.
- 김국희, 1992. 온라인 목록 탐색에 있어서 이용자 모형을 이용한 인터페이스의 설계에 관한 연구. 연세대학교 석사학위논문.
- 김명옥, 1989. 자료분류법. 구미무역(주)출판부.
- 김명옥, 1993. 서양의 자료분류법의 발달과정: 고대에서 해리스까지. 한국문헌정보학회지.
- 노정순, 유재옥 공저, 1996. 온라인정보탐색. 한국도서관협회.
- 사공복희, 1992. 온라인 목록의 이용자인터페이스에 관한 연구. 전남 대 현대사회과학연구.
- 유사라, 1996. 정보화 사회와 도서관 정보네트워크. 나남.
- 이용민, 1990. 온라인 목록 주제접근점 확장을 위한 분류표 지식베이스 설계에 관한 연구. 연세대학교 석사학위논문.
- 정영미, 1991. 최신정보검색론. 구미무역.
- 정명미, 김성은, 1997. WWW 탐색도구의 색인 및 탐색 기능 평가에 관한 연구. 한국문헌정보학회지.
- Ardian, A., Falcoz, F., Koch, T., Nielsen, M., Sandfaer, M., 1994. "Improving resource discovery and retrieval on the Internet." The Nordic WAIS/World Wide Web Project - Summary report. NORDINFO Nytt , 4, 13-28. <http://www.ub2.lu.se/W4/summary.html>
- Buchanan, B., 1979. Theory of library classification. London: Bingley.
- Buxton, A., 1990. "Computer searching of UDC numbers." Journal of Documentation, 46:3, 193-217.
- Chan, L.M., 1995. "Library of Congress Subject Headings: principles and application, 3rd ed." Littleton, Colo.: Libraries Unlimited.
- Cochrane, P.A., 1995. "New roles for classification in libraries and information networks." In: New roles for classification in libraries and information networks: reports from the Thirty-sixth Allerton Institute. Cataloging and

- Classification Quarterly, 21:2, 3-4.
- Courtois, et al., 1995. "Cool tools for searching the Web - a performance evaluation." Online, 19:6, 14-32
- Dahlberg, I., 1995. "The future of classification in libraries and networks: a theoretical point of view." In: New roles for classification in libraries and information networks: reports from the Thirtysixth Allerton Institute. Cataloging and Classification Quarterly, 21:2, 23-36.
- ✓Dodd, D.G., 1996. "Grass-roots cataloging and classification: food for thought from World Wide Web subject-oriented hierarchical lists." Library Resources & Technical Services, 40:3, 275-286.
- Foskett, A.C., 1973. The Universal Decimal Classification: the history, present status and future prospects of a large general classification scheme. London: Bingley.
- Guenther, R.S., 1996. "Automating the Library of Congress Classification Scheme: implementation of the USMARC Format for Classification Data." Cataloging & Classification Quarterly, 21:3/4, 177-203. INFOMINE, 1996, About INFOMINE  
<http://lib-www.ucr.edu/infomine/intro.html>
- Koch, T., 1996. "Searching the Web - Systematic overview over indexes."  
[http://www.ub2.lu.se/tk/websearch\\_\\_systemat.html](http://www.ub2.lu.se/tk/websearch__systemat.html).
- Koch, T., 1996. "The building and maintenance of robot based Internet search service: A review of current indexing and data collection methods."  
<http://www.ub2.lu.se/desire/radar/reports/D3.2>
- Koch, T., 1997. "The Role of classification schemes in Internet resource description and discovery."  
<http://www.ub2.lu.se/desire/radar/reports/D3.2>
- Koster, M., 1994. "ALIWEB: Archie-Like Indexing in the Web."

- <http://info.webcrawler.com/mak/projects/aliweb/paperwww94/>
- Langridge, D., 1991. Classifying knowledge. In: Meadows, A.J., (ed.), Knowledge and communication: essays on the information chain. London: Library Association Publishing, 1-18.
- Larson, R.R., 1992. "Experiments in automatic Library of Congress Classification." *Journal of the American Society for Information Science*, 43:2, 130-148.
- Marcella, R. and Newton, R., 1994. A new manual of classification, Aldershot: Gower.
- Markey, K., 1989. "Subject searching strategies for online catalogues through the Dewey Decimal Classification. In: Hildreth, C.R., (ed.), The online catalogue: developments and directions." London: Library Association, 61-83.
- McIlwaine, I. C., 1991. "UDC as a standard for subject control. In: McIlwaine, I.C., (ed.), Standards for the Exchange of Bibliographic Information": papers presented at a course held at University College London, 3-18 August 1990. London: Library Association.
- McIlwaine, I. C., 1995. "Preparing traditional classification for the future: Universal Decimal Classification." In: New roles for classification in libraries and information networks: reports from the Thirty-sixth Allerton Institute. *Cataloging and Classification Quarterly*, 21:2, 49-58.
- Mitchell, J., 1995. "DDC 21 and beyond: the Dewey Decimal Classification prepares for the future." In: New roles for classification in libraries and information networks: reports from the Thirty-sixth Allerton Institute. *Cataloging and Classification Quarterly*, 21 (2), 37-48.
- Mitchell, S., 1996. "Library of Congress Subject Headings as

- Subject Terminology in a Virtual Library: The INFOMINE Example"  
<http://lib-www.ucr.edu/pubs/postlcsh.html>
- Mitchell, J., et al., 1996. "Dewey decimal classification and relative index: devised by Melvil Dewey, 21st ed." 4 vv. Albany, NY.: Forest Press.
- Pinkerton, B., 1994. "Finding What People Want: Experiences with the WebCrawler."  
<http://info.webcrawler.com/bp/WWW94.html>
- Svenonius, E., 1983. Use of classification in online retrieval. Library Resources & Technical Services, 27:1, 76-80.
- Thompson, R., Shafer, K. and Vizine-Goetz, D., 1997. "Evaluating Dewey concepts as a knowledge base for automatic subject assignment." Dublin, Ohio: OCLC.  
[http://orc.rsch.oclc.org:6109/eval\\_dc.html](http://orc.rsch.oclc.org:6109/eval_dc.html)
- Venditto, G., 1996. "Search engine showdown." IW Labs tests seven Internet search tools, Internet World 7:5, 79-86
- Vizine-Goetz, D., 1996. "Using library classification schemes for Internet resources (Position Paper)." Proceedings of the OCLC Internet Cataloging Colloquium, San Antonio, Texas, January 19, 1996. Dublin, Ohio: OCLC.  
<http://www.oclc.org/oclc/man/colloq/v-g.htm>
- V Vizine-Goetz, D., 1996. "Online classification: implications for classifying and document-like object retrieval." Electronic version of a paper published in: Green, R., (ed.), Knowledge organisation and change: proceedings of the 4th international ISKO conference, 15-18 July 1996, Washington D.C. Frankfurt/M: INDEKS Verlag.  
<http://orc.rsch.oclc.org:6109/dvgisko.htm>
- Wallace, J. and Burden, P., 1995. "Toward a classification-based approach to resource discovery on the Web."

Wolverhampton: University of Wolverhampton, School of Computing and Information Technology.

<http://www.scit.wlv.ac.uk/wwlib/position.html>

Wyly, B., 1995. "What lies ahead for classification in information networks?" Report of a Panel Discussion. In: New roles for classification in libraries and information networks: reports from the Thirty-sixth Allerton Institute. Cataloging and Classification Quarterly, 21:2, 75-82.

## ABSTRACT

### A Study on the Classification Scheme of the Internet Search Engine

Kim, Young-Bo\*

The main purpose of this study is ① to settle and to analyze the classification of the Internet Search Engine comparatively, and ② to build the compatible model of Internet Search Engine classification in order to seek information on the Internet resources, specially in the branch of the Computers and Internet areas.

For this study, four Internet Search Engine (Excite, I-Detect, Simmany, Yahoo Korea!), Inspec Classification and two distionaries were used.

The major findings and result of analysis are summarized as follows:

1. The basis of the classification is the scope of topics, the system logic, the clearness, the efficiency.
2. The scope of topics is analyzed comparatively by the number of items from each Search Engine. In the result, Excite is the most superior of the four
3. The system logic is analyzed comparatively by the casuality, balance and consistency of the items from each Search Engine. Inthe result, Excite is the most superior of the four.
4. The clearness is analyzed comparatively by the clearness and accuracy of items, the recognition of the searchers. In the result,

---

\* Department of Library and Information Science Graduate School Sung Kyun Kwan University

Excite is the most superior of the four.

5. The efficiency is analyzed comparatively by the exactness of indexing and decreasing the effort of the searchers. In the result, Yahoo Korea! is the most superior of the four.
6. The compatible model of Internet Search Engine classification is established to uplift the scope of topics, the system logic, the clearness, and the efficiency. The model divides the area mainly based upon the topics and resources using 'bookmark' and 'shadow' concept.