

인터넷 정보서비스의 분류체계에 대한 비교연구 : 물리학을 중심으로

A Comparative Study on Classification Schemes of Internet Services

최 희 윤(Hee-Yoon Choi)*

목 차

- | | |
|---------------------|---------------------|
| 1. 서 론 | 5. 분류체계 비교분석 |
| 2. 인터넷 정보자원의 조직과 분류 | 5.1 계층구조 비교 |
| 3. 선행연구 개요 | 5.2 주제접근 방법 |
| 4. 인터넷 정보서비스의 분류체계 | 5.3 인터넷서비스의 탐색과정 비교 |
| 4.1 주제별 디렉토리 탐색엔진 | 5.4 비교결과 분석 |
| 4.2 문헌 분류체계의 적용 | 6. 결 론 |

초 록

인터넷 정보자원의 폭발적인 증가에 따라 이를 효율적으로 조직화하고 체계화하는 시스템의 중요성이 증가하고 있다. 이에 따라 주제접근을 용이하게 하고 검색 효율성을 높이는 도구로서 분류체계에 대한 관심이 커지고 있다. 본 고에서는 문헌 분류체계와 인터넷기반 분류체계의 계층구조와 접근방법을 구조적 측면과 검색사례를 통하여 조사하고 이에 대한 비교 분석을 통해 인터넷 환경에 적합한 분류체계의 구성방안을 살펴 보았다.

ABSTRACT

There is increasing importance of a system to reorganize explosive expansion of internet information resources efficiently; therefore, an increasing concern about classification system as an instrument for facilitating an access to a specific subject and improving efficiency in information retrieval.

Comparing the hierarchical structure and access methodology of internet-based classification system with those of library classification such as Dewey Decimal Classification through their structural aspects and retrieval process, this paper proposes the proper classification system in internet environment.

키워드 : 인터넷 정보서비스, 분류체계, DDC, 인터넷 탐색엔진, 주제별 디렉토리, Yahoo,
문헌 분류체계, 인터넷 기반 분류체계

* 포스코경영연구소 지적자산관리실장

■ 논문접수일 : 1998년 11월 25일

1. 서론

최근 들어 정보처리 기술의 발달과 인터넷 이용의 확산에 따라 인터넷 정보자원의 양이 폭발적으로 증가하고 있다. 이러한 인터넷의 잠재력은 컴퓨터에 입력된 정보자료의 급증과 더불어 이용자들이 직접 인터넷 정보의 제공과 신속한 접근이 가능하게 됨으로써 그 중요성이 더욱 커지고 있다. 특히 웹의 확산과 함께 정보제공의 주체는 주요 기관으로부터 개인으로까지 확대되었으며 하이퍼텍스트를 기반으로 한 대량의 정보검색이 가능하게 되었다. 즉 누구나 쉽게 인터넷상으로 정보를 출판할 수 있으며, 인터넷상에서 끊임없이 제공되는 엄청난 규모의 정보를 제공받을 수 있다. 그러나 인터넷을 통해 얻을 수 있는 정보의 양은 방대하지만 인터넷이 안고 있는 문제 또한 적지 않다. 특히 인터넷상에는 질적으로 통제되지 않는 가치없는 정보가 많다. 인터넷상의 정보는 인쇄출판물에 수반되는 편집상의 검토나 평가가 생략되기 때문에 신뢰성에 문제가 있으며, 그 속성상 비체계적이며 무조직적이어서 이용자들이 원하는 정보를 정확하게 찾는다는 것은 상용 데이터베이스와 비교하였을 때 매우 어려운 일이다.

따라서 방대한 인터넷 정보자원을 체계적으로 조직하여 이용자들이 필요로 하는 정보를 쉽고 빠르게 찾을 수 있도록 도와주는 다양한 검색도구들이 계속 개발되고 있다. 이것은 인터넷에서의 정보 검색의 어려움을 반영하는 증거라고 하겠다. 이 중에서 인터넷 정보자원을 조직하고 검색하는 효율적인 도구로서 네트워크 환경에서의 분류체계의 활용에 대한 관심이 증가하고 있다. 인터넷과 같은 온라인환경에서 분류가 검색효율성을 높이고 브라우저를 용이하게 하는 주제접근

도구로서 심각하게 검토되고 있는 것이다. 이미 Svonenius(1983)는 온라인환경에서도 분류체계를 이용하는 것이 적합율과 재현율을 높이고 이용자의 시간을 절약하며, 탐색어에 대한 문맥 및 브라우징 기능 제공, 상이한 언어간의 변환을 위한 메카니즘 제공 등을 실현하는 데 유용하므로, 용어간 상관관계를 계층적으로 보여 주며, 의미론적 브라우징에 도움이 되는 분류체계의 실제적인 적용이 필요하다고 주장하였다. 현재 인터넷을 통하여 다양한 서비스가 제공되고 있으며 이는 검색에 있어 분류체계의 잠재력을 재인식시키는 계기를 제공하고 있다.

본 고에서는 인터넷 정보자원의 조직과 검색을 위해 인터넷 기반 분류체계를 제공하는 주제별 디렉토리나 기존의 문헌분류체계를 적용한 인터넷 서비스의 현황을 조사하였다. 또한 물리학분야를 대상으로 Yahoo!와 DDC에 대한 분류체계의 계층구조와 주제접근방법은 물론 실제 이를 구현하는 인터넷서비스의 비교 분석을 통해 효율적인 분류체계의 구현방안을 제시하였다.

2. 인터넷 정보자원의 조직과 분류

모래사장에서 바늘을 찾는 것에 비유되는 인터넷 검색은 인터넷 정보자원이 얼마나 많은지, 부적합한 정보 또한 얼마나 많은지에 대한 단적인 표현이라고 볼 수 있다. Floridi(1996)는 인터넷상의 지식을 관리하는 기구가 없으므로 인해 분산적으로 생산되는 지식이 통합되지 못하고 단편적인 정보로 남게 된다는 점을 지적하고 있다. 따라서 인터넷은 질적으로 통제되지 않는 가치없는 정보의 증가, 신뢰성의 저하, 비체계적, 무조직적 정보자원의 확산이라는 문제점을 안고 있다. 이

문제를 해결하기 위해서는 온라인으로 생산되는 지식을 효과적으로 조직하고 체계화하는 작업이 필요하다. 이를 위한 기초 단계로 인터넷 정보자원에 대한 색인 데이터베이스, 분류, 목록, 주제별 디렉토리 등이 생산되고 있다. 이러한 탐색도구를 활용한 접근은 크게 다음의 두가지 유형으로 구분할 수 있다(정영미 1997, 242).

- 인터넷상의 정보자원들을 자동으로 색인하는 탐색엔진이나 디렉토리를 사용하는 것이다.
- 도서관 목록 작성과 같은 방식으로 인터넷자원들의 MARC 레코드를 작성하는것으로서 서지정보외에도 하이퍼링크를 통해 원문예의 접근도 가능하다. 현재 OCLC가 인터넷 정보자원의 편목 프로젝트를 수행하고 있으며, 미국 국회도서관은 LC의 WWW 서버와 Z39.50 게이트웨이를 통해 서지레코드와 인터넷정보자원을 연결하고 있다.

결국 인터넷 정보자원을 조직하는 방법은 색인과 구조화이며, 서비스 제공시에 이들은 통합화 경향을 보이게 된다. 색인된 정보자원은 키워드를 사용한 탐색의 대상이 되고, 분류개념을 이용해 구조화된 정보자원은 주제별 디렉토리로 조직되어 네비게이션의 대상이 된다고 할 수 있다(Denenberg, 1996).

3. 선행연구 개요

분류연구자들은 문헌 분류체계가 정보에 대한 주제접근을 향상시킬 수 있다는 것을 오래전부터 인식해 왔다. 앞서 언급한 Svonenius(1983)의 온라인 검색에서의 분류체계에 대한 연구 이래 Markey(Markey and Demeyer 1986)는 DDC 온라인 프로젝트에서 이용자의 주제접근을 개선

시키기 위한 분류체계의 구현방안을 최초로 제시하였다(Vizine-Goetz, 1996). 그러나 Markey의 선구적인 연구가 수행된지 10년이 지난 최근 들어서야 인터넷 이용의 증가와 함께 온라인환경에서 검색도구로서 분류의 유용성에 대한 검토가 심각하게 이루어지고 있다.

Svenonius의 연구를 이어 받아 Vizine-Goetz(1996)는 DDC와 LCC의 분류체계를탐색엔진 Yahoo의 분류체계와 비교, 분석하였는데 각 항목들의 용어와 구성에 포함된 수를 조사하여 항목간의 균형성과 적절성에 대한 문제를 밝혀내었다. 또한 Vizine-Goetz(1997) OCLC에서 이러한 분류체계를 적용한 사례를 레코드 형식과 구축방법에 대한 소개와 함께 제시하고 있다. Dahlberg(1995)는 네트워크 환경에서의 분류이론 적용에 관한 연구에서 LCC와 DDC를 선정하여 분류체계의 적합성 여부를 분석하였으며, 분류이론의 적용은 각 주제의 분석과 추적 및 계층 구분에 유용하며, 기존에 분류되어 있는 자료의 네트워크 재조직에도 필요한 것으로 보았다.

국내에서는 이명희(1997)가 주제별 디렉토리인 Yahoo와 키워드 검색엔진인 Alta Vista의 검색효율에 대한 비교연구를 수행하였으며, 김영보(1997)가 컴퓨터와 인터넷을 중심으로 인터넷 탐색엔진에 적용될 수 있는 분류체계의 모형을 구축하였다.

이와 같이 지금까지의 연구는 온라인목록의 분류체계 구축의 효율성 분석, 탐색엔진에 대한 검색 효율성의 평가, 기존 문헌 분류체계를 이용한 시스템 구축 등에 머무르고 있다. 인터넷 탐색엔진의 분류체계와 기존 문헌분류체계를 비교분석하여 인터넷 정보자원을 효율적으로 조직할 수 있는 최적의 분류체계와 서비스를 구축하려는 시도는 아직 이루어지지 않고 있다.

4. 인터넷 정보서비스의 분류체계

분류체계를 활용한 인터넷 정보자원의 조직화는 주제접근 도구로서 다음과 같은 특징을 갖추고 있다. 첫째, 정확한 주제접근과 주제어의 제시이다. 둘째, 검색용어의 문맥적인 의미를 제공한다. 동일 용어의 경우에도 관련주제를 제시함으로써 의미의 혼동을 감소시킬 수 있다. 셋째, 이용자에게 브라우징 기능을 제공한다. 넷째, 다국어간의 호환성을 제공할 수 있다. 이들은 다시 인터넷 기반의 분류체계를 제공하는 주제별 디렉토리, 검색엔진과 기존 문헌 분류체계를 활용한 서비스로 구분할 수 있다.

4.1 주제별 디렉토리 검색엔진

방대한 인터넷 정보자원중 필요한 정보를 찾을 수 도와 주는 인터넷 검색엔진은 그 유형에 따라 분류체계 제공형, 탐색/분류 통합형, 메타 검색엔진으로 나누어 볼 수 있다. 먼저 분류체계를 이용하는 주제별 디렉토리 검색엔진은 심마니, 정보탐정, Yahoo!, Magellan, Infoseek 등을 들 수 있으며, 이용자 편의 지향적이고 유연한 분류체계와 신속한 확장이 특징이다. 이러한 주제별 디렉토리는 주제별 목록을 제공하는 대표적인 서비스로서 특정적이지 않고 일반적이면서 막연한 개념을 대상으로 하는 탐색에서 보다 효과적이라는 장점이 있다(이명희 1997). 즉 주제별 디렉토리의 효용은 이용자가 찾고자 하는 정보의 개념을 논리적인 구조와 구체적인 주제명으로 제시해 주는 데 있다.

또한 탐색위주의 엔진은 Lycos, Alta Vista, HotBot, Infoseek, Excite 등을 들 수 있으며 이들은 데이터 양, 검색대상공간, 검색논리, 갱신의

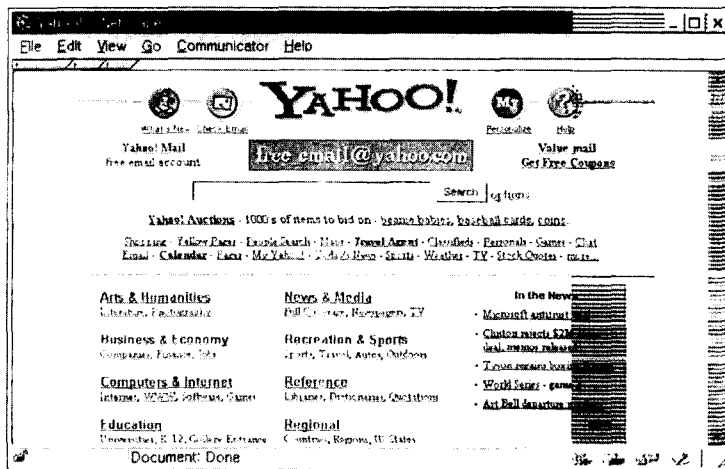
신속성, 탐색수행능력 등의 객관적인 평가요소를 갖추고 있다.

특히 자체적인 분류체계를 이용하여 인터넷 정보자원에 대한 효율적인 접근을 제공하는 서비스인 주제별 디렉토리는 Yahoo!에서 성공적으로 수행되었으며 많은 수의 검색엔진에서 이러한 주제 접근의 유용성을 인정하여 도입하고 있다. 이들은 상호참조, 전개상황 표시, 다양한 이용자에 대한 배려 등을 그 특징으로 하고 있으나 분류체계의 유용성에 대하여 평가할 수 있는 객관적인 척도는 많지 않다. 다음은 자체적인 인터넷기반 분류체계를 가지고 있는 주요 검색엔진의 분류체계 구성에 대한 조사결과이다.

1) Yahoo!의 분류체계 구성

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

Jerry Yang에 의해 1993년부터 개발되어온 Yahoo!는 현재 인터넷에서 가장 잘 알려진 사이트 중 하나이다. 이 사이트는 자체적으로 발전해 온 14개의 대분류를 통하여 인터넷 정보자원에 대한 접근을 제공하며 디렉토리 탐색기능을 제공하여 탐색어로도 직접 접근할 수 있도록 하고 있다. 또한 다양한 이벤트성 메뉴를 제공하여 주요한 내용이나 이벤트가 발생할 때 이에 대한 메뉴를 제공하여 주고 있다. 배열방법은 전 과정에 걸쳐 알파벳 순을 유지하고 있다. Yahoo!는 초기 화면을 대분류와 소분류로 구분하고 분류를 전개해 나갈 때마다 현재의 위치를 알려주고 있다. 야후의 구조는 상호참조적으로, 예를 들어 [Physics]항목의 [Chemical physics]는 실제로는 [Chemistry]에 존재하는 것으로 이용자는 [Physics]분야에 들어와서도 이에대한 접근점을 잃지 않을 수 있다. 즉, 최초 접근점을 어떠한 것으로 하더라도 동일한 결과에 이르게 되는 특징



〈그림 1〉 Yahoo!의 초기화면과 대분류 배열화면

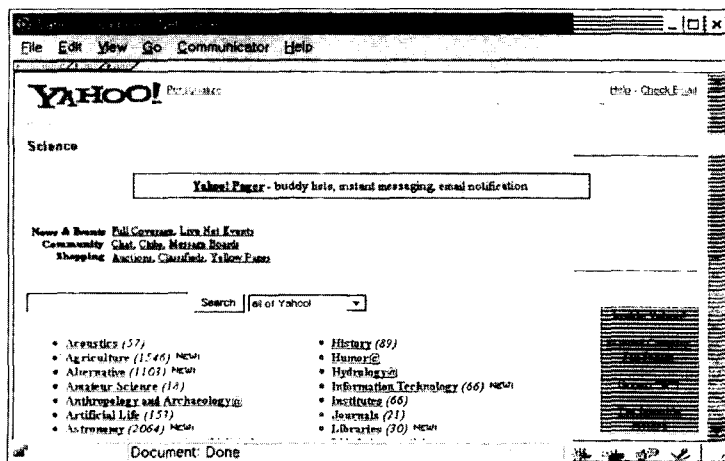
이라고 할 수 있다. 〈그림 1〉과 〈그림 2〉는 Yahoo!의 초기 분류체계와 물리학 분야의 검색사 레이다.

2) Infoseek의 분류체계 구성

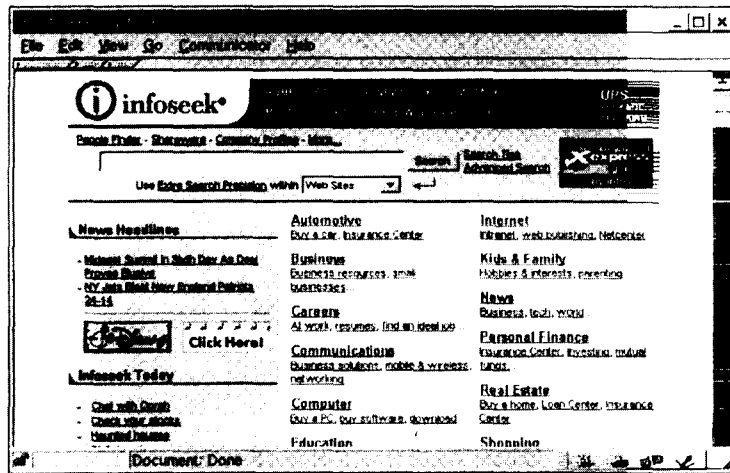
(<http://www.infoseek.com>)

Infoseek는 여러가지 다양한 정보자원을 직접

메뉴형태로 나타내어 주는 것이 특징이다. 초기 화면의 상단에 Stocks, News, Maps, People & Business, Reference, Yellow Page 등의 메뉴가 제공되어 있어 이들 정보를 원하는 이용자는 바로 접근할 수 있도록 되어 있다. 또한 화면 좌측에는 탐색어 입력을 위한 창이 제공되어 직접 탐색을 수행할 수 있다. 우측의 분류는 Yahoo!와



〈그림 2〉 Yahoo!의 물리학 분야 검색화면



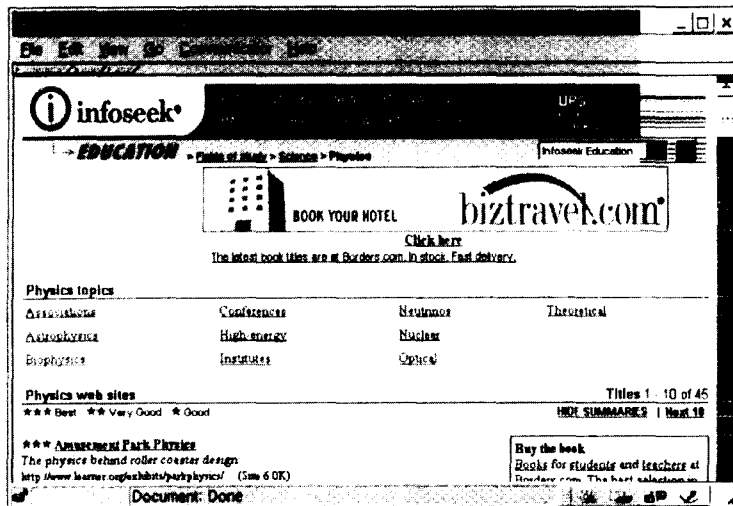
〈그림 3〉 Infoseek 초기화면의 대분류 배열화면

마찬가지로 알파벳 순으로 전개되어 있는데 Infoseek에서는 이를 채널이라는 이름으로 부르고 있다. 〈그림 3〉과 〈그림 4〉는 Infoseek의 초기 분류화면과 물리학 분야 소항목을 선택하였을 때 나타나는 인터넷 검색화면의 예이다.

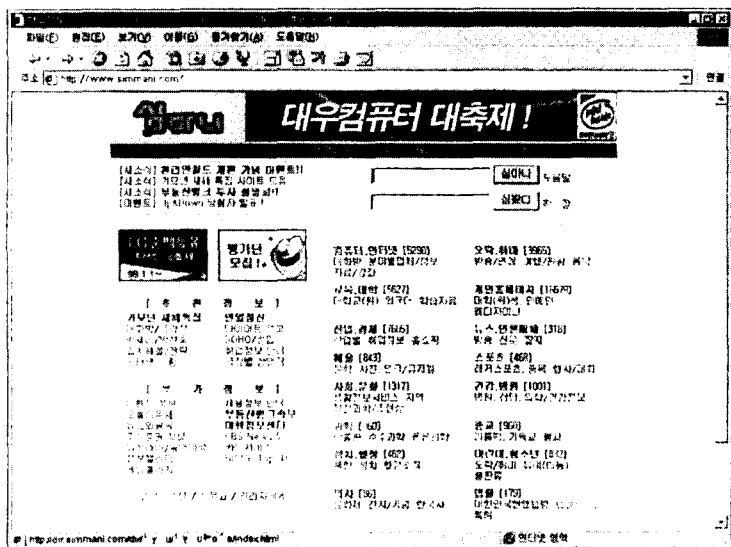
3) 삼마니의 분류체계 구성

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

삼마니는 한글을 처리할 수 있는 검색엔진이자 웹 디렉토리이다. 기초적인 검색기능과 더불어 유의어 및 한글발음의 확장검색을 지원하고 있으며 16개의 대분류를 통하여 다양한 정보자원을



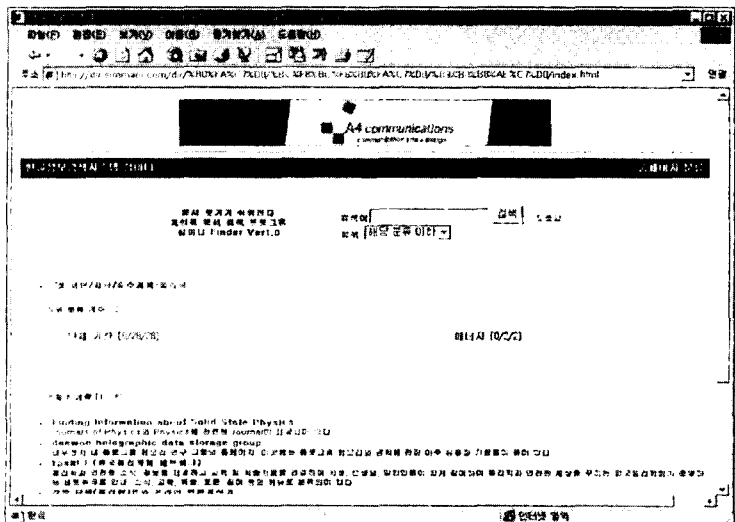
〈그림 4〉 Infoseek의 물리학 분야 검색화면



〈그림 5〉 심마니 초기화면의 대분류 배열화면

분류하고 있다. 초기화면의 좌측에는 추천정보와 부가정보를 통해 최신의 주요 정보원을 소개하여 이용자의 관심을 유도하고 있다. 또한 화면에서 직접 탐색어를 입력할 수 있는 창이 제공되어 직

접 탐색을 수행할 수 있으며 분류는 자체적으로 개발한 체계를 따르고 있다. 〈그림 5〉와 〈그림 6〉은 심마니의 초기 분류화면과 물리학 분야의 분류항목의 예이다.



〈그림 6〉 심마니의 물리학 분야 검색화면

〈표 1〉은 Yahoo!, Infoseek, 심마니의 대분류 체계와 각 하위분류에서의 개수를 비교한 결과이다. Yahoo!는 14개의 대분류와 464개의 하위분류를 보유하고 있으며, Infoseek는 17개의 대분류와 303개의 하위분류를, 심마니는 16개의 대분류와 213개의 하위분류를 보유하고 있다.

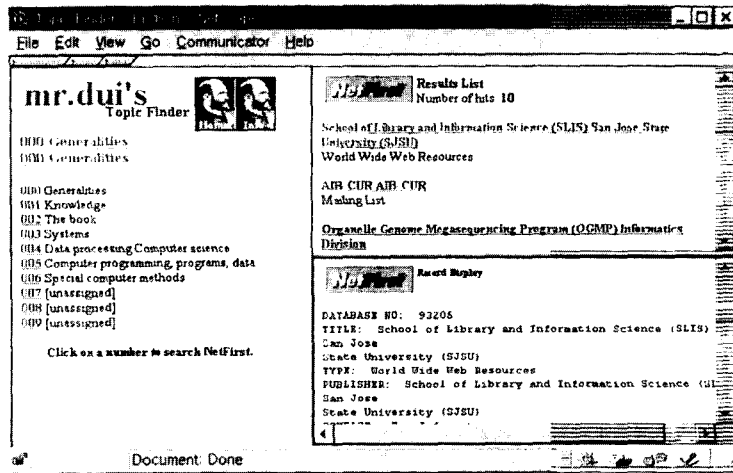
4.2 문헌 분류체계의 적용

최근에는 인터넷 정보자원의 증가와 함께 오랫동안 도서관분야에서 사용되어 온 방대한 문헌 분류체계인 DDC, UDC, LCC등을 인터넷 정보

자원의 주제분류에 이용하려는 다양한 시도가 이루어지고 있다. 이들은 주제에 대한 계층적인 접근이 특징이며 학술분야 위주의 서비스가 많다. 또한 이들 분류체계는 언어와 분야에 관계없이 총체적인 모든 장서를 배열하는 목적을 갖고 있다. 특히 DDC의 경우에는 지식을 구조화하고 배열하는데 계층적 구조를 사용하며, 기호를 사용하여 개념간 연산을 수행한다. 이러한 문헌분류체계의 인터넷 적용시도는 여러 분야에서 실험적으로 적용되고 있으며, 전체적으로는 DDC의 적용이 두드러진다(Koch, T., et. al 1997). 이들 서비스의 목표는 주제 위주의 브라우징 구조를

〈표 1〉 주제별 디렉토리의 분류체계

서비스명	Yahoo!	Infoseek	심마니
대분류의 개수	14 (464)	17 (303)	16 (213)
대분류의 내용 (하위분류수)	Arts & Humanities (25) Business & Economy (39) Computers & Internet (39) Education (34) Entertainment (32) Government (23) Health (44) News & Media (42) Recreation / Sports (24) Reference (37) Regional (3) Science (56) Social Science (39) Society & Culture (27)	Automotive (24) Business (24) Careers (13) Computer (25) Education (14) Entertainment (19) The Good Life (18) Health (23) Internet (20) Kids & Family (18) News (16) Personal Finance (12) Real Estate (12) Shopping (20) Sports (18) Travel (15) Women's (12)	컴퓨터, 인터넷 (12) 교육, 대학 (18) 산업, 경제 (13) 예술 (14) 사회, 문화 (24) 과학 (8) 정치, 행정 (7) 역사 (6) 오락, 취미 (44) 개인홈페이지 (5) 뉴스, 언론매체 (5) 스포츠 (13) 건강, 병원 (22) 종교 (6) 어린이, 청소년 (10) 법률 (6)



〈그림 7〉 Mr. Dui's Topic Finder의 검색화면
 (<http://www.oclc.org/oclc/fp/mrdui/eng1main.htm>)

제공하며 검색의 정확도를 향상시키고 검색결과를 주제별로 분리할 수 있도록 연산기능을 제공하는 등 다양한 접근방법을 통하여 시스템의 검색수행능력을 향상시키는 것이다.

4.2.1 DDC 적용 사례

DDC는 전형적인 계층분류법으로서 1876년 Dewey가 초판을 발행한 이래 현재 135개국 이상에서 사용되고 있으며, 30개 이상의 국어로 번역된 전세계적으로 가장 많이 활용되는 분류체계이다. 특히 21판은 4권으로 된 인쇄본뿐 아니라 "Dewey for Windows"(DFW)라는 윈도우즈버전과 웹버전인 "DDC Online"으로도 발간되었다(Hickey et. al 1997). DFW의 가장 큰 의의는 윈도우즈 버전으로 전자분류표가 나왔고 이용자에게 보다 많은 접근점을 쉽고 빠르게 제공하여 분류작업의 효율화를 가져올 수 있다는 점이다(정연경 1997, 94). DDC의 적용사례는 다음과 같다.

1) Mr. Dui's Topic Finder

(<http://www.oclc.org/oclc/fp/mrdui/mrdui.htm>)

OCLC의 DDC를 통한 다국어 지원 프로젝트인 이 사이트는 분류번호를 통해 주제에 접근하기 위한 목적으로 영어, 스페인어, 불어, 세가지 언어로 구성되어 있으며, 가장 체계적으로 분류가 이루어진 사이트이다. 화면은 3가지 프레임으로 구성되어 해설과 접근경로를 제공하고 있다. 〈그림 7〉은 001-Knowledge 분야의 목록과 그에 대한 해설이며 목록은 공식 명칭과 구성형태로 이루어져 있다.

화면 좌측의 분류번호를 따라 원하는 주제를 선택하다 보면 우측 상단 화면에 다음과 같은 주제명 디렉토리가 나타남을 알 수 있다.

한편 위에서 표시된 각 항목을 선택하면 다음과 같이 해당 정보자원의 상세한 서지사항이 우측 하단에 표시된다

Results List
 Number of hits: 10
 School of Library and Information Science
 (SLIS) San Jose State University (SJSU)
 World Wide Web Resources
 AIB-CUR AIB-CUR
 Mailing List
 Organelle Genome Megasequencing Program
 (OGMP) Informatics Division
 World Wide Web Resources

또한 여기에 사용된 데이터베이스는 OCLC의 인터넷 정보자원 데이터베이스인 NetFirst이다. OCLC에서는 이 데이터베이스를 이용하여 NetFirst라는 동명의 비공개 서비스도 운영하고 있다.

DATABASE NO: 93205
 TITLE: School of Library and Information Science (SLIS) San Jose State University (SJSU)TYPE: World Wide Web Resources
 PUBLISHER: School of Library and Information Science (SLIS) San Jose State University (SJSU)CONTACT: For Information office@wahoo.sjsu.edu
 ACCESS:
 (http:)http://witloof.sjsu.edu/homeslis.htmDO
 MAIN: edu Education
 SUMMARY: Features the School of Library and Information Science (SLIS) at San Jose State University (SJSU) in California.
 Includes information on the schedule of classes at the Fullerton and San Jose campuses as well as the faculty staff and alumni association. Discusses special projects the SLIS is involved in. Links to libraries library schools professional organizations in the information sciences and vendors of automated library systems with Web pages.
 LC SUBJECT: San Jose State University. School of Library and Information Science. Library schools — California — SanJose. Information science — Study and teaching — California — San Jose. Information science — Study and teaching — California — Fullerton. Library science — Study and teaching — California — Fullerton.
 DEWEY CLASS: 020.711 020.71 020.71 020.71

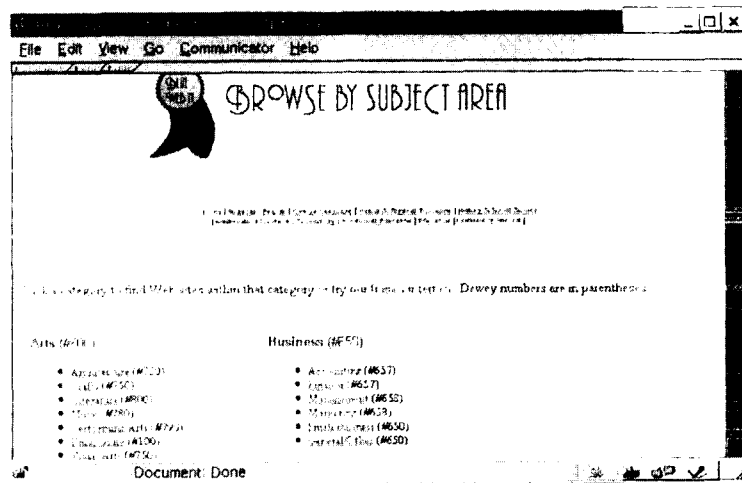
2) Blue Web'n Content Categories.
 (<http://www.kn.pacbell.com/wired/bluewebn/categories.html>)

이 사이트의 분류 전개 화면에서는 전체를 Arts, Business, English, Foreign Language, Health & Physical Education, History & Social Studies, Mathematics, Science, Technology, Vocational Education, Education, Community Interest의 12개 주요 분야로 나누어 각각을 DDC분류에 의해 제공하고 있다. 그중 Business(650)분야의 전개는 Business (#650), Accounting (#657), Finance (#657), Management (#658), Marketing (#658), Small Business (#650), General/Other (#650)으로 구성되며, 각 분류별로 추천사이트를 해설과 함께 제공하고 있다. 예를 들어 Business의 하위분류인 Management에서는 2개의 인터넷 사이트가 평가와 해설을 통해 제공되고 있다. <그림 8>은 주제 조회화면이다.

3) Internet Resources in Dewey Decimal Order with DDC Subjects.
 (<http://www.mcpl.lib.mo.us/dewey.htm>)

이 사이트에서는 모든 목록을 하나의 문서로서 관리하고 있어 전체가 하나의 주제별 목록이 만들어질 수 있다는 특징을 가지고 있다. 그러나 인터넷 주소를 명시하지 않아 출력했을 경우 목록으로서 기능하지 못한다는 단점을 가지고 있다.

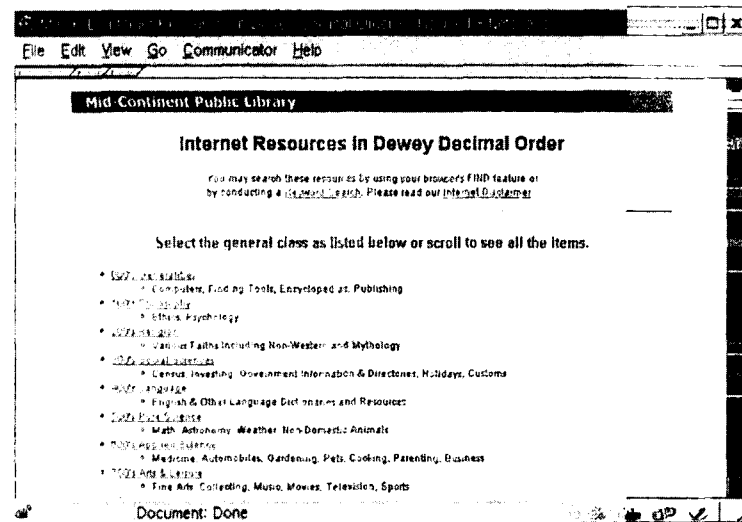
4) WEBrary(tm)
 (<http://www.webrary.org/re/weblinksmenu.html>)



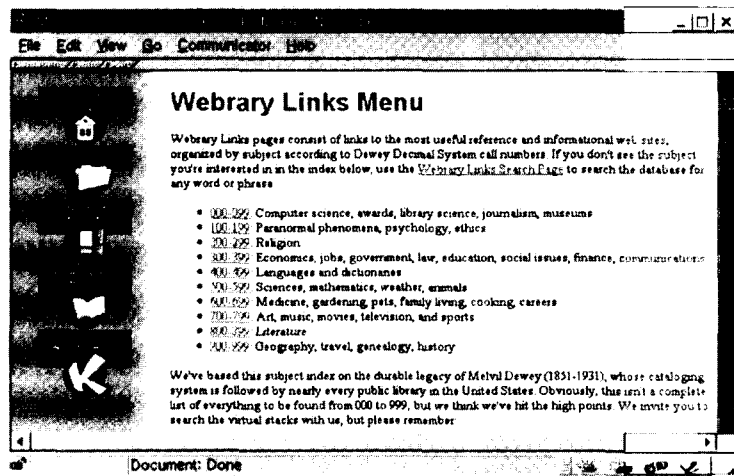
〈그림 8〉 Blue Web'n Content Categories의 주제조회 화면
(<http://www.kn.pacbell.com/wired/bluewebn/categories.html>)

일반적인 다른 사이트와 비슷한 형태를 취하고 있다. 모든 인터넷 정보자원에 대한 설명을 제공하고 있어 제목만으로 알기 어려운 사이트의 내용을 소개하고 있다. 특징적인 것은 DDC의 분

류체계를 따르고 있으면서도 분류번호에 대한 설명은 Yahoo! 등 인터넷 주제별 목록과 유사하다는 점이다. 이것은 분류체계와 이용자간의 간격을 좁히기 위한 시도라고 볼 수 있다.



〈그림 9〉 Internet Resources in Dewey Decimal Order with DDC Subjects의 주제조회 화면
(<http://www.mcpl.lib.mo.us/dewey.htm>)



〈그림 10〉 Webrary의 주제조회 화면 (<http://www.webrary.org/ref/weblinksmenu.html>)

5) WWLib Browse Interface

(<http://www.scit.wlv.ac.uk/wwlib/browse.htm>)

이 사이트는 각 주제분류를 들여쓰기를 통하여 매우 방대한 양으로 정리해 놓은 점을 특징으로 들 수 있다. 전체 보기를 원하는 경우 100번대 분류를 선택하면 되고 계속적인 전개를 원할 경우 각 항목을 골라 〈Examine〉 버튼을 눌러 전개해 나갈 수 있다는 점이 특징이다.

GERHARD는 쥐리히의 ETH 도서관에서 UDC를 확장하고 여러 언어를 지원하도록 개발된 자동화 프로젝트이며 독일어로 구성된 인터넷 정보 자원을 탐색하기 위한 목적을 가지고 있다. 분류 전개기능, 분류 탐색기능, 문서내용 탐색기능으로 구성되어 있으며 영어, 불어, 독일어를 지원하고 있다. 컴퓨터에 의해 수집된 100만개 이상의 문서를 UDC 엔트리에 따라 자동으로 분류한 실적을 가지고 있으며 계속 개발중인 프로젝트이며 〈그림 11〉은 Navigation의 화면 예이다.

4.2.2 UDC 적용 사례

UDC의 인터넷 정보자원에 대한 분류는 많은 사례가 발견되지는 않으나 특히 과학기술 분야에서 자동 분류를 연구하는 GERHARD의 사례가 흥미로우며 DDC와 마찬가지로 체계적인 분류와 목록방법을 연구하는 NISS 사례가 있다.

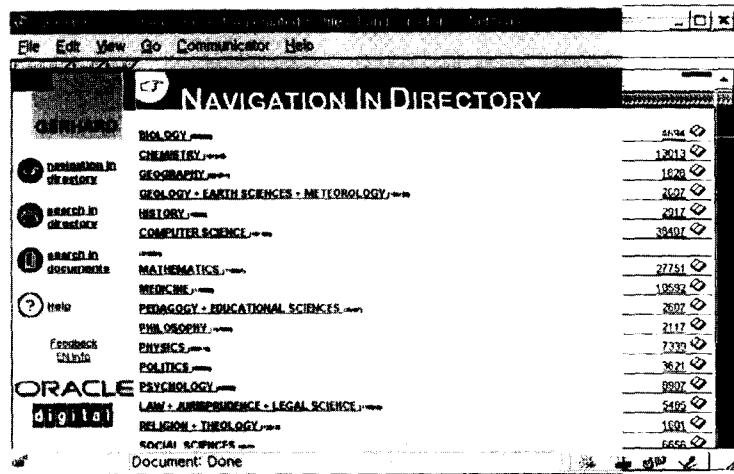
2) NISS, The NISS Directory of Networked Resources

(<http://www.niss.ac.uk/subject/index.html>)

NISS는 모든 주제분야에 대한 서비스를 제공하고 있으며 각 전문 주제분야별로 유사한 기능을 수행하는 디렉토리로의 하이퍼링크를 제공하고 있다는 점이 특징이다. 각각의 단계별로 분류번호별 목록과 그 단계에 해당하는 정보원과 접

1) GERHARD

(<http://gerhard.bis.uni-oldenburg.de/>)



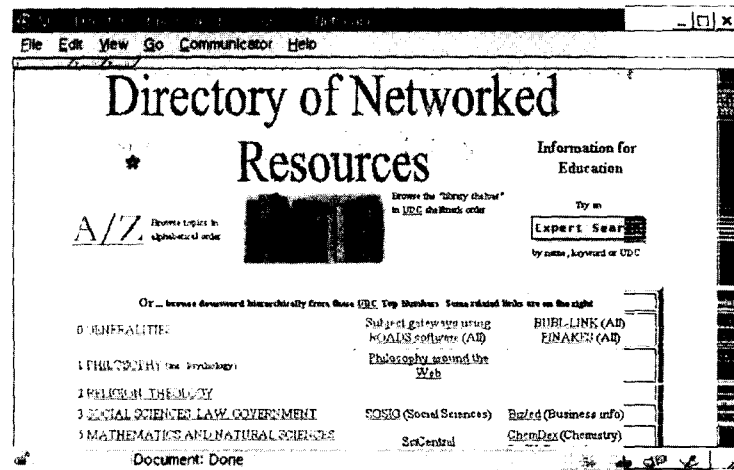
〈그림 11〉 GERHARD의 Directory Navigation
 (http://gerhard.bis.unioldenburg.de/gerold/owa/gerhard.create_index_html?form_language=1&form_timestamp=&form_language=1)

근방법, 그리고 해설을 포함하고 있는 것이 장점이라고 하겠다. 사이트에 대한 해설이 필드별로 체계적으로 되어 있어 별도의 링크를 제공한다는 점이 독특하나 목록 내에서 사이트의 해설을 볼 수 없다는 것은 불편한 점이다.

3) OMNI (Organising Medical Networked Information)

(http://www.omni.ac.uk/)

이 사이트는 의학분야의 인터넷 정보자원을 수집하여 서비스하고 있으며 초기에는 UDC의 체



〈그림 12〉 NISS Directory of Networked Resources의 분류전개 화면
 (www.niss.ac.uk/subject/index.html)

계를 따른 전개를 제공했으나 더이상 UDC를 사용하지 않고, NLM의 분류체계를 따른 60여개의 분류와 MeSH 주제명 표목, 그리고 알파벳 순서의 디렉토리를 통해 이용자의 선택권을 다양하게 제공하고 있다는 점이 특징이다.

4.2.3 LCC 적용 사례

미국 국회도서관 분류체계인 LCC는 이 도서관 장서에 맞도록 만들어진 것이기 때문에 문헌 분류이론에는 근거하고 있지 않다. 따라서 LCC의 경우에는 주로 주제명으로 접근하는 시도가 눈에 띄었으며 포괄적인 분류에 대한 연구에서는 다른 분류체계에 비하여 그 구축량이 적었다. 그러나 OMNI의 사례에서 볼 수 있듯이 MeSH 주제명 표목을 활용한 사례도 이용자 인터페이스를 위한 다양한 방법으로 제시될 수 있다(Koch et al, 1997).

1) CyberStacks

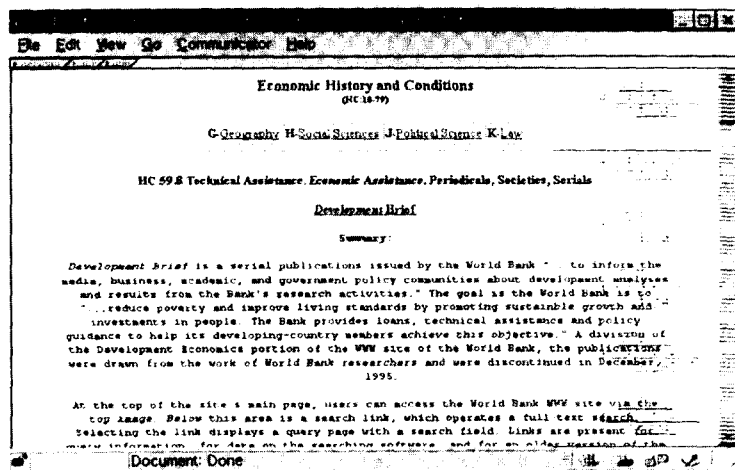
(<http://www.public.iastate.edu/~CYBERSTACKS/hybrow-all.htm>)

현재 G(Geology), H(Social Sciences), Q(Science), R(Medicine), S(Agriculture), T(Technology) 분야만이 제공되고 있으며 아직 완성된 단계는 아닌, 개발단계로 볼 수 있다. 그러나 수집된 정보 자원에 대해서는 상세한 해설을 하고 있는 점이 특징이다.

2) The WWW Virtual Library

(<http://www.w3.org/pub/DataSources/bySubject/LibraryOfCongress.html>)

W3 Consortium에서 운영하는 사이트로 계층적 주제분류를 사용한 매우 상세한 주제별 디렉토리를 제공하고 있다. 즉 이용자로서는 분류기호를 모르는 상태에서 서비스에서 제공하는 주제명만을 선택하여 필요한 정보자원을 고르게 되며, 분류체계는 내부적으로만 사용되고 있음을 알 수 있다. 실제로는 일단계 하위주제로 세분되



〈그림 13〉 CyberStacks의 인터넷 정보자원 디렉토리 설명단계 (http://www.public.iastate.edu/~CYBERSTACKS/hyb_hc_1.htm)

는 매우 구체적인 140여개의 주제범주를 사용하는 분류체계 및 LCC 분류체계의 두 체계에 따라 인터넷 자원을 조직하고 있다.

3) Electronic Collections Online (<http://www.oclc.org/oclc/eco/service.htm>)

인터넷상에서 전자잡지를 제공하는 대표적인 프로젝트로 OCLC는 1996년 Electronic Journals Online 서비스를 통해 여러 출판사와 학협회가 제공하는 전자잡지를 웹을 통해 유료로 제공하다가 중단하고 1997년 중반부터는 대신 Electronic Collection Online Service를 제공하고 있다. 전자잡지들은 LCC 분류에 따라 분류되어 있으며 모두 9개의 주제 및 다수의 하위 주제 아래 열거되어 있다.

5. 분류체계 비교분석

본 조사에서는 물리학 분야를 비교영역으로 하여 문헌분류체계인 DDC와 인터넷기반 분류체계인 Yahoo!를 비교하였다. 물리학 분야를 선정한 이유는 이 분야가 근대과학의 초기부터 많은 성과를 보유하고 있으며 최근에 이르기까지 많은 이론적, 실험적 분과학문과 그 응용사례를 포함하고 있기 때문이다. 분류체계에 있어 기존에 세워진 학문체계와 새롭게 추가해야 할 항목을 분배할 필요가 있을 때 적용하는 방법론은 실제 구현에 있어 중요한 역할을 하므로 비교영역으로는 물리학 분야를 대상으로 하였다.

인터넷에서는 Yahoo! 이외에도 수많은 디렉토리 서비스가 존재하고 있음에도 Yahoo!를 선정한 이유는 기타 서비스들은 단순한 분류체계

로 인해 DDC와의 비교 변별력이 부족했기 때문이다. 앞서 살펴본 바와 같이 국내의 대표적인 인터넷 디렉토리 분류 서비스인 심마니(<http://www.simmani.com>)의 경우 물리학 분야의 분류체계는 <그림 6>에서와 같이 매우 단순함을 알 수 있다. 심마니의 경우 물리학 분야의 주제분류는 [단체, 기관]과 [에너지]의 2가지밖에 없다. 이런 상태에서는 학문적인 주제구분을 비교할 수 있는 상황이 아니므로 심마니의 분류체계는 비교분석에서 제외하였다.

또한 Yahoo!와 유사한 검색엔진인 Infoseek나 Excite 등은 Yahoo!와 같이 체계적인 분류체계가 아닌 기초적인 수준의 분류항목만을 구분한 것을 앞서 살펴본 <그림 4>를 통해 확인할 수 있었다. 즉 Infoseek나 Excite 등은 비교를 수행할 정도의 분류체계가 세분되어 있지 않았으며, 결국 Yahoo!만이 비교분석을 수행할 수 있는 정도의 분류체계를 제공하였다.

또 하나의 문제는 DDC가 이미 확립된 분류체계인 반면 인터넷에는 출판되고 표준도구가 되어 널리 사용되는 분류체계라고 할 만한 것이 존재하지 않는 데서 발생한다. 인터넷에서는 각각의 서비스가 이용자에 따라 가장 적합하다고 결정한 분류결과만이 존재할 따름이다. 따라서 인터넷 분야에서는 가장 보편적으로 사용되는 Yahoo!의 분류체계를 그 대상으로 하였다. 즉 인터넷 세계에서 가장 많이 사용되는 서비스라면 이용자에게 인터넷 자원을 대상으로 하여 분류의 장점을 제시할 수 있는 것으로 볼 수 있다.

또한 비교분석에서는 순수한 분류체계와 그 결과인 서비스를 별도로 구분하여 조사하였다. 이는 서비스의 인터페이스 활용성으로 인한 편향을 방지하기 위한 것이다. 또한 인터넷 서비스상에서 분류체계의 구현에 필요한 요소를 찾기 위해

실질적인 인터페이스 개발요소도 함께 분석해 보았다.

5.1 계층구조 비교

주제의 계층구조는 분류체계의 기초가 된다. 주어진 자료가 그 주제특징에 따라서 어떠한 경로로 최종항목에 이르는지를 알 수 있기 때문이다. 따라서 어떤 분야에 관심있는 이용자가 분류의 계층구조에 따라서 어떻게 그 항목에 이를 수 있는지를 살펴보는 것이 실용적인 관점이라 하겠다. 예를 들어 물리학 분야의 열역학을 찾고 있는

이용자에게는 최상위 분류에서부터 원하는 항목으로 유도해 주는 용어의 선택과 계층 설계가 중요할 것이다.

5.1.1 주류항목

주류(Main class)는 이용자가 가장 먼저 접하는 영역으로 이곳에서 자신이 찾고자 하는 항목을 결정할 수 있도록 구성되어야 한다. 먼저 <표 2>에서 DDC와 Yahoo의 주분류 계층을 알아보고 이를 비교하기로 한다.

분류에 사용된 용어를 중심으로 비교했을 때 Yahoo의 경우가 보다 구체적인 용어를 사용하였

<표 2> DDC와 Yahoo의 주류항목 비교

DDC Numbers	Yahoo! Categories
000 Generalities	Reference, Computers & Internet, News & Media
100 Philosophy & psychology	
200 Religion	
300 Social sciences	Social Science, Society & Culture, Business & Economy, Education, Government
400 Language	
500 Natural sciences & mathematics	Science
600 Technology (Applied sciences)	Health
700 The arts Fine and decorative arts	Arts & Humanities, Entertainment, Recreation & Sports
800 Literature & rhetoric	
900 Geography & history	Regional

음을 알 수 있다. 예를 들어 Generalities 보다는 Reference, Computer, News & Media 등은 이용자가 쉽게 접하는 용어이다.

한편 계층구조를 중심으로 비교했을 때 DDC의 경우 Generalities를 통하여 Reference, Computer, Media 등의 공통점을 하나의 단계로 구분하고 있다. Yahoo에서는 이러한 단계에 대한 구분이 불명확하다. 예를 들어 Business & Economy와 Social Science가 최상위 분류에 함께 나타나고 있는데 이렇듯 Business를 포함시킴으로써 사회과학이 경제학보다 상위의 개념임에도 불구하고 Yahoo의 이용자는 Economy = Social Science라는 구조를 받아들여야 한다. 이러한 점이 주류항목에서 일어날 수 있는 계층상의 오류라고 할 수 있다.

또한 DDC와 Yahoo는 주제집중도가 상대적으로 판이하였다. Yahoo에서는 최상위계층에 등장하지 않는 Language나 Literature 분야가 DDC에서는 주류에 항목을 점유하고 있다. 또한 DDC에서는 Social Science로 대표되는 사회과학 분야의 영역이 Yahoo에서는 Social Science, Society & Culture, Business & Economy, Education, Government 등으로 등장하고 있다. 이러한 현상은 기호의 사용과 관련된다고 하겠다. 분류기호가 결정되어 있는 DDC에서는 정해진 분류번호를 지속적으로 유지해야 하는 반면, 기호를 사용하지 않고 용어만을 사용하는 Yahoo에서는 비교적 자유롭게 변화요소를 수용할 수 있기 때문이다. 또한 주제집중도의 차이는 그 분류대상이 되는 정보자원의 차이에서도 비롯될 수 있다. 인터넷에서 구할 수 있는 많은 정보자원이 Language나 Literature 분야에 있지 않다면 인터넷 서비스로서는 그들을 최상위 계층에 올려둘 필요가 없기 때문이다.

5.1.2 계층구조

DDC의 물리학분야 최상위 단계(530/539)와 비교하여 Yahoo의 같은 계층에서는 Fluid, Magnetism, Optics의 3개 항목만이 일치하였다. 또한 DDC는 원리적인 분류항목배열을 Yahoo는 구체적인 학문배열을 보여주었다.

DDC와 Yahoo의 계층적 차이는 물리학 분야를 세부적으로 살펴보면 두드러진다. 1차 항목(DDC 10개, Yahoo 23개)의 경우 두 체계간에 밀접한 관계를 가지고 있는 항목은 Fluid Dynamics(mechanics), Magnetism, Optics(Optical Engineering)의 3개 항목이다. 그리고 DDC에서 하위분류에 속해있는 항목이 Yahoo에서는 첫단계에 나타나는 경우가 4개 항목이 된다(Theories and mathematical physics → Theoretical Physics, Temperature → Temperature, Atomic and nuclear physics → Atomic Physics/Nuclear Physics, Particle Physics → High-Energy and Particle Physics) 이것은 Yahoo의 계층구조가 매우 구체적이고 실질적인 용어로 이루어져 있는 반면 DDC의 경우에는 이론적인 구조를 중심으로 설계되어 있기 때문이다.

또한 DDC에서는 최근에 중요성이 높아진 많은 최신분야가 539(Modern Physics)에 밀집될 가능성을 보여주고 있다. 이러한 특성은 서가탐색에 있어서는 물리학의 발전 순서에 따른 열람의 이점을 제공하지만 실제적으로는 분류기호가 복잡해지며 이용자는 원하는 주제를 찾아 더 깊은 계층까지 들어가야 하며, 수집자료의 분류항목간 정보자원의 편중이 심화되는 것이 단점이라 하겠다.

한편 Yahoo는 연결구조의 특징을 나타내었다. Physics 항목 하위에 있는 많은 분류는 실제로는

〈표 3〉 DDC와 Yahoo의 물리학 분야 비교

DDC Numbers	Yahoo! Categories
530 Physics 530.1 Theories and mathematical physics 530.4 States of matter 530.8 Measurement	Science Physics Astrophysics Atomic Physics Biophysics Chaos Chemical Physics Condensed Matter Crystallography Fluid Dynamics High-Energy and particle Physics Holograms Lasers Magnetism Mechanics Nuclear Physics Optical Engineering Photonics Plasma Physics Quantum Chemistry, Computing Relativity Superconductivity Temperature Theoretical Physics X-Ray
531 Classical mechanics, Solid mechanics 531.11 Dynamics	
532 Fluid mechanics, Liquid mechanics 532.2 Hydrostatics 532.5 Hydrodynamics	
533 Pneumatics (Gas mechanics)	
534 Sound and related variations	
535 Light and paraphotic phenomena, Optics	
536 Heat 536.2 Heat transfer 536.5 Temperature 536.7 Thermodynamics	
537 Electricity and Electronics	
538 Magnetism	
539 Modern Physics 539.1 Structure of matter 539.6 Molecular physics 539.7 Atomic and nuclear physics 539.72 Particle physics	

Physics의 것이 아니며 다른 분류항목과 연결되어 있는 구조이다(분류항목 다음에 '@' 표시가 나타나는 항목이 그 예임). 즉 많은 부분이 Physics의 직접적인 하위분류가 아니라는 점은 이용자로 하여금 계층관계를 혼동케할 우려가 있

다. 물론 이러한 특징은 인터페이스와 접근성이 라는 평가기준에서는 장점으로 작용할 수도 있다. 〈표 3〉은 DDC와 Yahoo의 물리학 분야의 계층구조 비교를 나타내고 있다.

〈표 3〉에서 나타난 Yahoo의 하위분류 가운데

다른 계통에서 유래된 항목은 다음과 같다.

- Astrophysics ← Home : Science : Astronomy :
- Chaos ← Home : Science : Mathematics :
- Chemical Physics ← Home : Science : Chemistry :
- Fluid Dynamics ← Home : Science : Engineering : Mechanical Engineering :
- Holograms ← Home : Science : Physics : Lasers :
- Optical Engineering ← Home : Science : Engineering :
- Quantum chemistry ← Home : Science : Chemistry: Computational chemistry :

Yahoo는 독립된 분과학문을 알파벳순으로 열거하는 배치와 다양한 분야에서의 접근점을 제공하고 있으나 DDC에서 제공되는 물리학의 주요 분야를 빠뜨리는 경향이 있음도 발견되었다. Yahoo에서는 Sound, Variation, Waves 등 물리학의 기초영역인 진동과 파동에 대한 항목은 발견되지 않아 앞으로 보다 다양한 분야의 확충이 필요한 것으로 조사되었다.

5.2 주제접근 방법

주제접근 방법은 기호법과 같이 분류체계의 기본은 아니며 보조도구에 가깝다. 그러나 DDC의 상관색인(Relative Index) 또는 Yahoo!의 연결구조 등은 주어진 주제명으로 가장 적합한 분류항목을 찾아갈 수 있도록 도와주는 역할을 수행한다. 따라서 이 기능은 계층을 확립하는 분류의 일차적인 기능은 아니더라도 보편적인 분류체계

에는 반드시 제공되어야 할 핵심기능이라고 할 수 있다.

또한 최근의 학문간의 협력 또는 학제적 연구의 증가 추세는 이용자들이 다양한 시작점과 경로를 통하여 분류체계에 접근할 수 있어야 함을 알려주고 있다. 따라서 분석방법은 DDC에서 사용된 물리학 분야의 용어가 Yahoo에서 어떠한 경로를 거쳐 몇단계만에 도달하게 되는가를 Yahoo의 검색기능을 통해 알아보는 방법을 사용하였다. 이러한 방법은 주제접근성을 분류체계의 계층구조와는 분리하여 서비스에서의 구현에 초점을 맞추고 있는 것이므로 모든 카테고리를 검색을 통해 확인할 수 있는 Yahoo의 경우를 중심으로 살펴보았다.

1) Physics

최상위 계층에서 Business and Economy, Entertainment, Government, Health, Regional, Science, Society and Culture의 7가지의 분야를 통한 접근이 가능하다. 그 내용은 다음과 같다.

- * Business and Economy : Companies
→ 4개 하위분류 포함
- * Entertainment : Physics Humor
- * Government : US : Government
→ 4개 하위분류 포함
- * Health : Environmental Health : Radiation and Health Physics
- * Regional(국별, 지역별)
→ 110여개 하위분류 포함
- * Science
- * Society and Culture(노벨물리학상)

2) Relativity theory(상대성이론)

3가지 분야를 통한 접근이 가능하며 대부분의 경우 Regional의 분류카테고리가 가장 많다.

- * Science : Physics : Relativity
(학문으로서의 상대성이론)
- * Business and Economy : Companies : Books : Science : Titles : Physics : Relativity
(상대성이론에 관한 논문, 저서 등)
- * Regional : Countries : UK : Science : Physical : Relativity (영국에서의 상대성이론)

3) Dynamics(역학)

Business and Economy, Government, Regional, Science의 4가지 분야별 접근이 가능하다.

- * Business and Economy : Companies
- * Government : US Government
- * Regional
- * Science : Engineering : Mechanical Engineering

첫 번째의 Business and Economy와 같은 경우 회사명에 Dynamics가 포함되어 있는 카테고리도 존재하였다. 특히 회사 분류에 자주 나타나며, 정보자원과 분류항목을 혼동할 소지도 가지고 있다.

위의 조사에서 확인된 바와 같이 실제 제공되는 서비스로서 Yahoo는 다양한 접근방법을 보여주고 있다. 특히 계층구조 조사에서는 다른 분야로 생각되는 Regional 분야의 각 지역별 연구기관이나 교육기관을 카테고리 수록한 것을 확인할 수 있었다. 또한 조사 (2)에서 확인할 수 있었던 바와 같이 상대성 원리를 다루는 과학출판사의 타이틀을 제공하는 등 독특한 접근경로를 제공하고 있었다. 그러나 DDC의 많은 용어에 있어

Yahoo에서는 전혀 분류항목으로 다루지 않은 분야가 많았으며 공통되는 항목 역시 많지 않았다.

Yahoo에서의 주제접근 조사를 통하여 계층으로만 조사했을 때 알 수 없었던 Business 분야와 Regional 분야의 카테고리 많음을 확인할 수 있었다. 이러한 현상은 Yahoo의 분류체계가 다양한 용어와 학문분야를 포괄하지 못하면서 비교적 정보자원의 수집이 용이한 기업과 지역 분야에 치중한다는 인상을 주고 있다. 그러나 이론적인 분류체계와는 별도로 Yahoo는 실제로 사용되는 분류목록이라는 관점에서 보면 인터넷 상에 나타나 있지 않은 정보원에 대한 분류항목은 없는 것이 타당하다. 따라서 나타나지 않은 분류항목은 적어도 Yahoo 서비스에서는 찾아내지 못한 인터넷 정보자원의 항목으로 이해할 수 있다. 이러한 추론을 진행시켜 보면 인터넷에서 제공되는 분류 서비스의 내용은 그들이 찾아낼 수 있는 인터넷 정보자원의 수량과의 함수관계에 있을 수도 있다. 이러한 가정에 대한 해답은 향후에 광범위한 연구와 조사가 이루어진 후에 가능할 것이다.

한편 이러한 컴퓨터화된 접근방법과는 달리 DDC의 상관색인은 특정한 기호가 사용되는 상황을 분명하게 하기 위하여 상위분류를 표현하는 충분한 용어를 분류항목명 다음에 부가하게 된다. 이러한 방법은 하위분류로부터 기초하는데 예를 들면 539.12의 Molecular structure은 "Molecular structure : Structure of matter 539.12" 와 같은 표현을 갖게 된다. <표 4>는 이러한 530 항목을 중심으로 한 조사 결과를 나타내고 있다.

5.3 인터넷 서비스의 탐색과정 비교

지금까지 물리학분야를 대상으로 Yahoo와

DDC 자체의 계층구조와 주제접근에 대하여 비교분석해 보았다. 다음으로 실제 DDC를 적용한 인터넷 서비스의 탐색과정을 Yahoo와 비교해 볼 수 있다. 앞서 언급한 바와 같이 DDC는 다양한 분야의 인터넷서비스에 적용되고 있으나 아직은 소규모의 로컬서비스 수준이며, 이중 대다수는 문헌분류체계의 적용 가능성에 대한 실험적인 서비스로 볼 수 있다. 즉 Yahoo와 같이 많은 정보자원을 보유하고 조직화한 대규모 비교대상 서비스의 확보가 어려운 점이 있다. 이중에서 다른 인터넷디렉토리에 비해 정보자원의 선별성 및 다양성, 고품질 서지레코드, LC 주제명표목과 DDC의 제공, 편집자의 초록 제공 등에서 우수한 것으로 평가되는 OCLC의 인터넷 정보자원 디렉토리인 NetFirst를 선정하였다. NetFirst는 WWW사이트, ListServ 메일링리스트, Usenet Newsgroups, Anonymous FTP 사이트를 대상으로 1997년 말 기준으로 90,000건의 레코드를 포함하고 있으며, 매달 8,000건의 레코드가 추가되고 있다.

5.3.1 NetFirst의 탐색과정

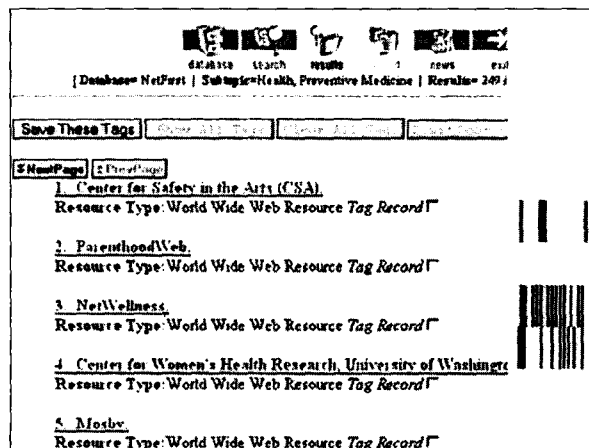
NetFirst는 DDC에 기반하고 있지만 분류기호 자체는 사용하지 않고 각 항목별로 이용자들에게 친숙한 용어를 사용하고 있다. 즉 DDC의 계층적 구조를 용어로 변환하고, 이에 따라 주제접근을 제공한다.

NetFirst에서는 주제명의 카테고리과 주제를 선택하고 하위주제를 선택하는 과정에서 14,000여개의 전체 데이터베이스에서 필요한 레코드를 효율적으로 필터링한다. 즉 주제접근은 필요한 정보를 효율적으로 필터링해주는 결과를 보인다. 예를 들어 <그림 15>에서 <그림 17>은 "Tourism, Travel"이라는 주제와 "Health, Medicine"이라는 주제의 복합검색에 대한 예를 보인 것이다.

(1) "Health, Medicine"을 주제명에서 선택한 예비검색 (249건)

(2) 과정 (1)에 "Tourism, Travel"의 키워드 검색을 결합 (25건)

(3) 단순히 키워드로만 검색 데이터베이스를 검색 (143건)
DDC Numbers



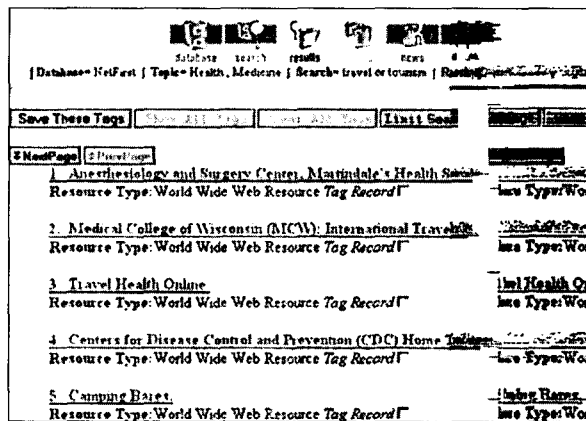
<그림 15> DDC에서 주제명으로서의 접근 및 검색과정
(1) "Health, Medicine"을 주제명에서 선택한 예비검색 (249건)

〈표 4〉 DDC에 기초한 Yahoo의 물리학분야 주제접근 비교

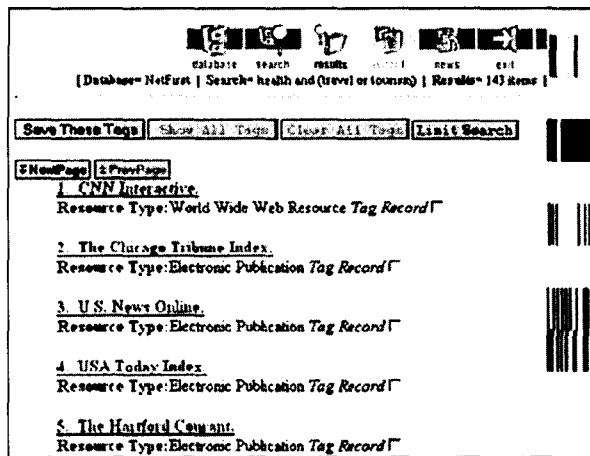
DDC Numbers	Yahoo Categories
530 Physics	<p>Yahoo Categories에서는 아래와 같이 다른 주제분야의 하위 분류체계로서의 물리학분야가 나타나 있다. 이들은 Physics라는 항목 아래 나타나 있지 않은 것이 대부분이며 결국 다른 분류항목으로부터 전개된 물리학 분야의 분류항목이다.</p> <p>Business and Economy : Companies 6개 하위분류 (Home : Business and Economy : Companies : Scientific로부터 전개된 물리학 응용분야의 회사를 할당하고 있다. 즉 물리학 응용기업은 물리학으로부터 전개되는 것이 아닌 기업의 과학분야로부터 전개되고 있으나 Yahoo의 물리학 항목에도 디스플레이 되는 것이 특징이다)</p> <p>Entertainment : Physics Humor (물리학 유머의 경우에도 Entertainment의 하위분류인 물리학 유머라는 항목을 물리학 분야에서 볼 수 있다)</p> <p>Government : US Government 4개 하위분류 (물리학과 관련된 정부기관이 아니라 정부기관 중 물리학과 관련된 기관으로 분류한 것으로서 이를 물리학 항목내에서 볼 수 있게 디스플레이 해 주고 있다)</p> <p>Health : Environment Health : Radiation & Health Physics (건강, 환경 항목으로부터 방사능과 이와 관련된 보건 물리학 분야를 표시하고 있다)</p> <p>Regional 110개 하위분류(각 국가, 지역명) (물리학 분야의 지역구분이 아니라 지역별 물리학 관련 연구, 단체, 기관 등을 전개하는 형식으로 표시되고 있다)</p> <p>Science : Physics와 다른 분야와의 연결 (이 경우는 Astrophysics나 Chemical Physics와 같이 다른 분야와 결합된 형태의 분과학문을 지정하고 있다)</p> <p>Society and Culture (노벨물리학상) (Home : Society and Culture : Issues and Causes : Philanthropy : Organizations : Grant-Making Foundations : Nobel Foundation과 같은 분류를 통하여 이루어진 노벨 물리학상을 물리학 분야에서 표시하고 있다)</p>
530.1 Theories & math physics	<p>Science : Physics : Theoretical Physics : Theories Science : Physics : Mechanics : Gravity : Theories</p>
530.11 Relativity theory	<p>Science : Physics : Relativity (학문으로서의 이론) Business and Economy : Companies : Book : Science : Titles : Physics : Relativity (상대성 이론에 대한 저서) Regional : Countries : UK : Science : Physics : Relatively</p>

〈표 4〉 DDC에 기초한 Yahoo의 물리학분야 주제접근 비교(계속)

DDC Numbers	Yahoo Categories
530.12 Quantum mechanics	Science : Physics : High-Energy and Particle Physics (양자역학) Science : Physics : Quantum computation (양자계산) Science : Physics : Quantum Teleportation (양자이동) Regional Countries Swizerland : Cantons : Zureig (스위스의 연구개발 프로젝트)
530.124 Wave mechanism	Science : Physics : High-Energy and particle Physics : Wave-Particle Duality (파동입자) Science : Engineering : Mechanical Engineering : Dynamics/Vibrations/Waves (엔지니어링)
530.13 Statistical mechanics	없음
530.133 Quantum statistics	없음
530.15 Mathematical physics	별도의 학문분야가 존재하지 않음 Regional하위분류에 학위과정 소개관련 1개 항목 존재
530.8 Measurement	없음 (과학분야에서 Science : Measurements and Units 분야에서 다름)



〈그림 16〉 DDC에서 주제명으로서의 접근 및 검색과정
(2) 과정 (1)에 "Tourism, Travel"의 키워드 검색을 결합 (25건)



〈그림 17〉 DDC에서 주제명으로서의 접근 및 검색과정
 (3) 단순히 키워드로만 검색 데이터베이스를 검색 (143건)
 (Health, Home, Technology→Health and Medicine→Health Preventive Medicine)

따라서 주제에 보다 근접한 검색결과를 얻기 위해 두번째 경우와 같이 분류와 키워드 검색을 결합하는 방법이 유용함을 알 수 있다. 이러한 사례를 통해 NetFirst의 주제접근에 대한 유용성을 구체적으로 확인할 수 있다(Vizine-Goetz 1997).

5.3.2 Yahoo의 특성 및 탐색과정

Yahoo는 각 분류전개에 있어 실제적인 주제분류와 공통항목분류가 결합되어 있다. 즉 Yahoo는 주제분류 + 공통항목의 결합이라는 두가지 축을 중심으로 통합된 시스템이다. Yahoo의 용어중복도는 공통부분에 110개 용어가 있으며 이들의 중복도는 56.70%로서 매우 높은 편이다. 반면 순수 주제별 분류의 중복도는 9.96%로서 2차전개 후 총 281개의 분류체계에서 253개의 독립된 주제명을 가지고 있었다. 하위분류로 평균 21개의 항목을 포함하고 있어 DDC에 비하여 많은 편이다.

Yahoo에서는 Health 이후에 Medicine이라는

하위분류항목이 있으므로 이를 검색대상으로 하였으며 검색건수는 Yahoo내에서 제공되는 정보 자료의 수를 참조하였다. 또한 "Health, Travel" 항목으로부터 Medicine을 키워드 검색하는 과정도 추가하였다. 동일한 결과에 이르는 여러 방법에서 빠진 경로를 추적하기 위해서이다.

(1-1) "Health, Medicine"을 주제명에서 선택한 예비검색 (4655건)

(1-2) 과정 (1-1)에 "Travel"의 키워드 검색을 결합 (22건)

(1-3) 단순히 키워드로만 검색 데이터베이스를 검색 (56건)

(2-1) "Health, Travel"을 주제명에서 선택한 예비검색 (20건)

(2-2) 과정 (2-1)에 "Medicine"의 키워드 검색을 결합 (6건)

(1-3) 단순히 키워드로만 검색 데이터베이스를 검색 (56건)

지금까지의 검색결과는 Health - Medicine/Travel과 같이 Yahoo의 2차분류까지를 적용한 것이다. 따라서 NetFirst와 마찬가지로 Yahoo에서도 주류인 Health만을 대상으로 검색해 보기로 한다.

(3-1) "Health"을 주제명에서 선택한 예비검색 (17,368건)

(3-2) 과정 (3-1)에 "Travel"의 키워드 검색을 결합 (84건)

(3-3) 단순히 키워드로만(medicine+travel) 검색 데이터베이스를 검색 (56건)

위의 조사를 통해 NetFirst와 유사한 결과가 나타남이 확인되었다. 그러나 NetFirst와 동일한 방법을 취한 세 번째 조사의 경우 예비 검색영역이 너무 광범위한 이유로 오히려 재현율을 떨어뜨릴 수 있음을 알 수 있었다. 조사방법 (1)과 (2)는 주제영역을 Medicine과 Travel로 구분하여 접근하였는데 이에 대한 결과 역시 상이하여 주제별 분류에 따른 오류가 나타날 수 있음을 확인하였다. 따라서 향후 서비스의 분류체계의 평가시 접근 순서에 의한 결과의 차이도 유의한 평가기준으로 검토되어야 할 것이다. 서비스별로 사용할 개념을 다양하게 조합하여 시도하는 것보다 정확한 결과를 얻음을 알 수 있다. 또한 검색에 사용한 용어는 여러 가지 실험을 거쳐 가장 높은 재현율을 확보하는 Travel과 Medicine으로 한정하였음을 밝혀둔다.

5.4 비교결과 분석

지금까지 문헌분류체계와 인터넷기반 분류체계의 특성을 DDC와 Yahoo를 대상으로 하여 계층

구조, 접근방법, 인터넷서비스의 탐색과정 등 3단계를 통해 비교 분석하였다. 먼저 계층구조에 있어서는 DDC에서는 분류체계의 고정성이 갖는 취약점과 Yahoo의 경우 항목간의 계층할당에 있어서의 오류가능성과 주요한 학문영역을 빠뜨릴 가능성이 발견되었다. 이들 문제점은 향후 서비스 설계에 있어 우선적인 고려대상이 되어야 할 것이다. 또한 접근방법에 있어서는 DDC의 각 분류항목에 대해 Yahoo에서 검색할 수 없는 많은 항목이 발견되었다. 이 현상은 분류체계가 사용한 용어의 차이가 직접적인 원인으로 파악되며 이를 극복하기 위해서는 주제어에 대한 많은 연구와 수용이 중요하다고 하겠다. 마지막으로 실제로 구현된 서비스를 대상으로 탐색과정을 비교하였다. 주제분류와 키워드검색을 통한 접근이 보다 정확한 결과를 넘으로써 분류체계가 검색에 있어 유용한 도구가 될 수 있음을 확인하였다. 다만 접근 순서에 의해 결과가 달라질 수 있음이 밝혀졌으며 앞으로 이를 처리하기 위한 기준이 연구되어야 할 것이다.

따라서 효율적인 인터넷 서비스의 분류체계는 이용자가 쉽게 찾아갈 수 있는 접근점과 이의 순서를 우선적으로 고려해야 하며, 분류항목의 논리적 구성과 현상을 적절하게 결합할 필요가 있다. 또한 분류기호는 공통항목과 순수주제로 구분하여 설계되어야 한다.

6. 결 론

인터넷 상에서 지식의 효율적인 전달과 이용을 위해 지식을 어떤 체계로 조직할 것인가에 대한 관심이 점점 높아지고 있다. 이러한 상황에서 인터넷 정보자원을 조직하고 검색하는 도구로서 검

색 효율성을 높이고 주제접근을 용이하게 할 수 있는 분류체계의 유용성에 대한 검토가 필요하다. 실제로 다양한 인터넷 탐색엔진과 문헌 데이터베이스에서 분류개념을 적용하여 인터넷 정보 자원의 조직과 접근의 효율성을 도모하고 있다. 이를 위해 본 고에서는 인터넷 정보자원의 조직과 검색을 위한 분류개념의 적용현황을 기존 문헌분류체계와 인터넷기반 분류체계를 중심으로 비교분석해 보았다. 또한 이들 분류체계의 비교평가를 통하여 인터넷에서 구현되어야 할 이상적인 분류체계의 구현방안을 살펴보았다.

분류체계의 계층구조에 있어서는 DDC 분류체계의 고정성이 갖는 취약점과 Yahoo의 계층오류 가능성을 고려하여 유연성있게 설계되어야 한다. 또한 주제 접근방법에 있어서도 관련 용어의 상이점을 고려하여 가장 적합한 분류항목의 용어와

이를 지원하는 주제용어집을 구축하여야 할 것이다. 또한 조사를 통하여 분류체계가 방대한 정보원으로부터 정확한 정보를 필터링할 수 있는 정보조직 방법이며, 어떠한 방법으로 접근하여도 동일한 수준의 해답을 제공할 수 있도록 분류체계가 설계되어야 함을 확인하였다.

인터넷 정보자원의 효율적인 조직과 검색을 위해서는 이용자의 정보이용행태를 면밀히 분석·파악하는 연구가 우선적으로 수행되어야 할 것이며, 이 경우 이용자의 정보요구 분석결과를 토대로 원하는 주제에까지 직접적으로 접근하는 경로설계에 대한 연구가 필요하다. 아직까지 국내에서 인터넷 정보자원을 대상으로 하는 분류기준에 대한 연구는 초기단계에 있으며, 최적의 분류모형을 설계하기 위한 지속적인 연구가 수행되어야 할 것이다.

참 고 문 헌

- 김영보. 1997. "인터넷 탐색엔진의 분류체계에 관한 연구: 컴퓨터, 인터넷분야를 중심으로," 성균관대학교 대학원 석사학위논문.
- 이명희. 1997. "네트워크 데이터베이스에서의 주제별 디렉토리 및 키워드 탐색엔진의 탐색효율에 관한 탐색적 연구," 한국문헌정보학회지, 3(2) : 177-197.
- 정연경. 1996. "Electronic Dewey의 이용에 관한 연구," 제3회 한국정보관리학회 학술대회 논문집, 63-66.
- 정연경. 1996. "듀이십진분류표의 인쇄형과 전자형 비교 및 이용에 관한 연구" 정보관리학회지 13(2) : 97-119.
- 정영미. 1997, 지식구조론. 서울: 한국도서관협회.
- Alexander, Jan & Tate, Marsha. 1996. "Teaching Critical Evaluation Skills for World Wide Web Resources," *Computers in Libraries* 16(10) : 49-55. <http://www.science.widener.edu/~withers/webeval.htm>.
- Bendig, Mark W. 1994. "Mr. Dui's Topic Finder," <http://www.oclc.org/oclc/research/publications/review96/dui.htm>
- Chan, L. M. 1995. "Classification, Present and Future," *Cataloging & Classification*

- Quarterly*, 21(2) : 5-18.
- Cochrane, P. A. 1995. "New Roles for Classification in Libraries and Information Networks," *Cataloging & Classification Quarterly*, 21(2) : 3-4.
- Dahlberg, Ingtraut. 1995 "The Future of Classification in Libraries and Networks, a Theoretical Point of View," *Cataloging & Classification Quarterly*, 21(2) : 23-36.
- Denenberg, R. 1996. "Structuring and Indexing the Internet,"
<http://lcweb.loc.gov/z3950/agency/italy.htm>.
- Floridi, L. 1996. "The Internet: Which Future for Organised Knowledge, Frankenstein or Pygmalion? Part 1 & Part 2," *The Library Journal*, 14(1) : 43-52.
- Hickey, C., et. al 1997. "Knowledge Organization Research,"
http://www.oclc.org:5047/~vizine/home/ko_main.html.
- Janes, J. W. and Rosenfeld, L. B. 1996. "Networked Information Retrieval and Organization : Issues and Questions," *JASIS*, 47(9) : 711-715.
- Kent, R. E. and Neuss, C. 1994. "Creating a Web Analysis and Visualization Environment,"
<http://www.informatik.th-darmstadt.de/~neuss/www2/wwwf94.html>.
- Koch, T., et. al 1997. "The Role of Classification Schemes in Internet Resource Description and Discovery,"
<http://www.ub2.lu.se/desire/radar/reports/D3.2.3/>.
- Library of Congress. 1997. "CD-ROMs from the Cataloging Distribution Service,"
<http://lcweb.loc.gov/cds/cdroms1.html>.
- OCLC, 1998. "Dewey Decimal Classification,"
<http://www.oclc.org/oclc/fp/index.htm>.
- Svenonius, E. 1983. "Use of Classification in Online Retrieval," *Library Resources & Technical Services*, 27(1) : 76-81.
- Vizine-Goetz, D. 1997. "OCLC Investigates Using Classification Tools to Organize Internet Data,"
<http://www.oclc.org/oclc/new/n226/research.htm>.
- Vizine-Goetz, D. 1996. "Using Library Classification Schemes for Internet Resources," Proceedings of the OCLC Internet Cataloging Colloquium,
<http://www.oclc.org/oclc/man/colloq/v-g.htm>.
- Vizine-Goetz, D. 1996. "Classification Research at OCLC,"
<http://www.oclc.org/oclc/research/publications/review96/class.htm>.