

인터넷 정보자원의 조직을 위한 분류체계에 관한 비교분석

Comparative Analysis of Classification Schemes for Internet Resources

최희윤, 포스코경영연구소

Choi, Heeyoon, POSCO Research Institute

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. 본고에서는 인터넷 서비스의 분류체계 응용사례중 주제별 디렉토리의 인터넷 기반 분류체계와 기존 문헌 분류체계를 비교 분석함으로써 인터넷 환경에 적합한 분류체계의 구성방안을 제시하였다.

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 정보에 대한 제공과 신속한 접근이 가능하게 됨으로써 그 중요성이 더욱 커지고 있다. 그러나 인터넷을 통해 얻을 수 있는 정보의 양은 방대하지만 인터넷이 안고 있는 문제 또한 적지 않다. 특히 인터넷상에는 질적으로 통제되지 않는 가치없는 정보가 많아 신뢰성에 문제가 있으며, 그 속성상 비체계적이며 무조직적이어서 이용자들이 원하는 정보를 정확하게 찾는다는 것은 상용 데이터베이스와 비교하였을 때 매우 어려운 일이다. 이를 해결하기 위해서는 인터넷상에서 온라인으로 생산되는 지식을 효과적으로 조직화하고 체계화하는 작업이 필요하다.

따라서 인터넷 정보자원을 조직하고 검색하는 효율적인 도구로서 네트워크 환경에서의 분류체계의 활용에 대한 새로운 관심이 증가하고

있다. 인터넷과 같은 온라인환경에서 분류가 검색효율성을 높이고 브라우징을 용이하게 하는 주제접근 도구로서 심각하게 검토되고 있는 것이다. 본고에서는 인터넷 정보자원의 효율적인 조직을 위해 인터넷 기반 분류체계를 제공하는 주제별 디렉토리나 기존의 문헌분류체계를 적용한 인터넷서비스를 조사하고 인터넷 서비스 상에서 분류 및 서비스체계의 특징을 비교 분석하여 효율적인 분류체계의 구성방안을 제시하고자 하였다.

2. 선행연구 개요

분류연구자들은 문헌 분류체계가 정보에 대한 주제접근을 향상시킬 수 있다는 것을 오래 전부터 인식해 왔다. 이미 Svonenius(1983)는 온라인환경에서도 분류체계를 이용하는 것이 적합함과 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 브

라우징에 도움이 되는 분류체계의 실제적인 적용이 필요하다고 주장하였다. Markey(Mark and Demeyer 1986)는 DDC 온라인 프로젝트에서 이용자의 주제접근을 개선시키기 위한 분류체계의 구현방안을 최초로 제시하였다(Vizine-Goetz, 1996). 그러나 Markey의 선구적 연구 이후 최근 들어서 인터넷 이용의 증가와 함께 온라인환경에서 검색도구로서 분류의 유용성에 대한 검토가 심각하게 이루어지고 있다. 또한 인터넷 탐색엔진이 널리 보급되면서 이들 탐색엔진 사이의 분류체계와 성능을 비교 분석하는 연구가 함께 수행되기 시작하였다.

Svenonius의 연구를 이어받아 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 네트워크상 재조직에도 필요한 것으로 보았다. 또한 Wallis와 Burden(1995)은 웹 자원에 대한 분류체계 기반 탐색의 분석을 통해 DDC를 응용한 WWlib과 같은 주제별 디렉토리 탐색엔진이 텍스트기반 탐색에서 특히 우수한 성능을 보임을 밝히고, 이를 통해 Automated Classification Engineer, ACE) 시스템을 제안하였다. 이들은 분류체계에 기반한 주제별 디렉토리 탐색엔진 설계가 많은 탐색엔진에 적용되어야 한다고 보았다.

국내에서는 이명희(1997)가 주제별 디렉토리인 Yahoo!와 키워드 검색엔진인 Alta Vista의 검색효율에 대한 비교연구를 수행하였으며, 김

영보(1997)가 컴퓨터와 인터넷분야를 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.

3. 인터넷 정보자원의 분류체계

3.1 주제별 디렉토리 탐색엔진

인터넷 탐색엔진은 그 유형에 따라 분류체계 제공형, 탐색/분류 통합형, 메타 탐색 엔진으로 나누어 볼 수 있다. 이중 분류체계를 이용하는 주제별 디렉토리 탐색엔진은 배열된 주제에 따라 이용자가 찾고자 하는 정보의 개념을 논리적인 전개와 구체적인 주제명으로 제시해 주므로 막연한 개념을 대상으로 하는 탐색에서 효과적이다. 또한 키워드 탐색엔진은 데이터 양과 갱신의 신속성 등의 평가요소를 갖추고 있으며 메타정보 탐색엔진은 여러가지 탐색엔진에 동시에 접속하여 질의결과를 회수하는 방식을 취하는 특징을 가지고 있다. 특히 자체적인 분류체계를 이용하여 인터넷 정보자원에 대한 효율적인 접근을 제공하는 서비스인 주제별 디렉토리는 Yahoo! 등에서 성공적으로 수행되었으며 Infoseek, Hotbot 등 많은 수의 탐색엔진에서 이러한 주제 접근의 유용성을 인정하여 도입하고 있다. 각 시스템의 조사 결과 Yahoo!는 14개의 대분류와 464개의 소분류를 보유하고 있으며, Infoseek는 17개의 대분류와 303개의 소분류를, HotBot은 13개의 대분류와 156개의 소분류를 보유한 분류체계를 유지하고 있다.

3.2 문헌 분류체계의 활용

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

자원의 주제분류에 이용하려는 다양한 시도가 이루어지고 있다. 이러한 시도는 여러 분야에서 실험적으로 적용되고 있으며, 전체적으로 DDC의 적용이 두드러진다(Koch, T., et. al 1997).

3.2.1 DDC 적용사례

◆ Mr. Dui's Topic Finder

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

OCLC의 DDC 적용 프로젝트인 이 사이트는 NetFirst와 함께 가장 체계적으로 분류가 이루어진 사이트이다(Koch, T. et al 1997). 화면은 3프레임으로 구성되어, 화면 좌측은 분류가 지속적으로 이루어지고 화면 우측은 최종 선택시 내용이 디스플레이 된다. 우측 화면은 다시상·하단으로 구분, 상단은 인터넷 정보자원의 목록이 나타나고, 하단은 체계적인 목록 체계에 따라 해설과 접근방법을 제시하고 있다.

3.2.2 UDC 적용사례

◆ GERHARD

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

GERHARD는 쥐리히의 ETH 도서관에서 UDC를 확장하고 여러 언어를 지원하도록 개발된 프로젝트이며 독일어 인터넷 정보 자원을 탐색하기 위한 목적을 가지고 있다. 전개/탐색 기능, 문서 열람기능으로 구성되어 있으며 컴퓨터에 의하여 수집된 100만개 이상의 문서를 UDC 엔트리에 따라 자동으로 분류한 실적을 가지고 있으며 계속 개발중인 프로젝트이다.

3.2.3 LCC 적용 사례

◆ The WWW Virtual Library

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

계층적 주제분류를 사용한 매우 상세한 주제별 디렉토리를 제공하고 있다. 분류기호를 몰라도 서비스에서 제공하는 주제명을 선택하여

필요한 정보자원을 고르게 되며, 구체적인 140여개 주제범주의 분류체계 및 LCC 분류체계에 따라 인터넷 자원을 조직하고 있다.

3.3 비교 분석

위의 사례조사를 참고하여 인터넷기반 분류체계를 사용한 주제별 디렉토리나 문헌 분류체계를 활용한 서비스를 비교 분석해 보기로 한다. 대표적인 주제별 디렉토리로는 Yahoo!를 들 수 있으며 문헌 분류체계의 활용은 DDC를 적용한 OCLC의 NetFirst가 대표적이라고 할 수 있다.

3.3.1 주제 집중도

학술적 측면에서 DDC와 Yahoo!를 비교한 결과 Yahoo!에서는 철학, 종교, 언어, 문학 부문이 대분류에서 제외되어 있다. 반면에 총류, 사회과학, 예술 분야에는 여러 가지 주제명으로 대분류가 3개 이상씩 집중되어 있어 DDC와 차이가 있음을 알 수 있다. <표 1>은 DDC와 Yahoo!의 주제 집중도를 비교한 결과이다.

<표 1> DDC와 Yahoo!의 주제집중도 비교

| DDC | Yahoo! |
|---------------------------------------|--|
| 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 |

또한 Yahoo!는 특히 주제분류와 공통분야의 분류를 통합하여 제공하고 있다. 즉 Yahoo!의 자생적 분류 체계는 [기본 주제분류 + 공통 항목의 결합]이라는 두가지 축을 하나의 통합시스템으로 발전시킨 것으로 볼 수 있다.

3.3.2 계량적 분석

분류체계의 항목수가 적으면 빠르게 전개할 수 있는 대신에 네비게이션이 복잡해지는 단점이 있으며, 항목수가 많아지면 이용자가 원하는 주제를 일별하는데 문제가 발생한다. 일반적으로 인터넷 기반 분류체계의 평균 항목수는 문헌분류체계와 비교하여 많았다. <표 2>는 Yahoo!와 Hotbot, Infoseek, Topic Finder의 항목수를 비교한 것이다.

<표 2> Yahoo!, Hotbot, Infoseek, Topic Finder의 항목수 비교

| 구분 | Yahoo! | HotBot | Infoseek | Topic Finder |
|-------------|---------------|---------------|----------------|--------------------|
| 초화면 주제분류수 | 13 | 13 | 17 | 10 |
| 초화면 접근점 | 135 | 104 | 133 | 11 |
| 2차분류 전체갯수 | 475 | 157 | 292 | 100 (3차분류 = 1,000) |
| 주제의 개수 | 281 | 128 | 260 | 100 |
| 공통항목 개수 | 194 | 29 | 32 | - |
| 전체항목수(중복제외) | 337 | 137 | 260 | 100 |
| 중복항목수 | 138 | 20 | 30 | - |
| 평균분류갯수 | 21 | 12 | 17 | 10 |
| 배열방식 | 알파벳순 (일관됨) | 알파벳순 (일관됨) | 알파벳순 (예외있음) | DDC 분류순 |

또한 인터넷 기반 분류체계는 중복되는 공통항목을 포함하는 것이 특징이다. 이를 Yahoo!를 통해 확인한 결과가 <표 3>에 나타나 있다.

<표 3> Yahoo!의 분류항목 분석

| 대분류명 | 전체 | 공통 | 주제별 |
|------------------------|--------|--------|-------|
| Arts & Humanities | 25 | 14 | 11 |
| Business | 39 | 21 | 18 |
| Computers and Internet | 39 | 15 | 24 |
| Education | 34 | 11 | 23 |
| Entertainment | 31 | 13 | 18 |
| Government | 23 | 11 | 12 |
| Health | 44 | 11 | 33 |
| News & Media | 42 | 12 | 30 |
| Recreation & Sports | 22 | 6 | 16 |
| Reference | 37 | 32 | 5 |
| Regional | 17 | 7 | 10 |
| Science | 55 | 20 | 35 |
| Social Science | 39 | 12 | 27 |
| Society & Culture | 28 | 9 | 19 |
| 계 | 475 | 194 | 281 |
| 용어중복제거후 | 337 | 84 | 253 |
| 용어중복 건수 | 138 | 110 | 28 |
| 용어의 중복도 | 29.05% | 56.70% | 9.96% |

<표 3>에서와 같이 공통부분의 194개에 해당하는 용어가 각 분류에 적용되고 있다. 이들의 중복도는 56.70%로서 매우 높은 편이다. 반면 순수 주제별 분류의 중복도는 9.96%로서 2차전개후 총 281개의 분류체계에서 253개의 독립된 주제명을 가지고 있으며 분류당 평균 21

개의 소분류를 유지하고 있다. 이러한 점은 새로운 주제발생시 기존의 분류에 추가하는 형태의 확장이 이루어졌음을 보여주고 있다. 인터넷 정보자원을 대상으로 하고 있는 Yahoo!는 이러한 정책을 사용함으로써 분류체계의 유연성을 확보하는 장점과 동시에 논리성이 약화되는 문제점을 드러내게 되었다.

3.3.3 분류 전개과정과 이용자 인터페이스

한편 기존 문헌분류체계와 인터넷기반 분류체계의 검색효율성 측면을 비교하기 위하여 의학 분야에 대한 분류체계를 중심으로 DDC의 NetFirst와 Yahoo!의 전개과정을 살펴보았다. NetFirst 데이터베이스는 DDC의 계층적 구조를 사용하여 인터넷 정보자원에 대한 주제접근을 제공하는 기능을 갖고 있다. "tourism, travel"이라는 주제와 "health, medicine"이라는 복합주제에 대한 검색의 경우 주제명과 키워드 탐색을 결합한 방법이 주제에 근접함을 보여주고 있다. (Vizine-Goetz 1997).

이러한 과정을 Yahoo!는 이용자 인터페이스를 중시하며 접근한다. 100여개에 가까운 소항목은 이미 분류체계라기 보다는 주제명의 알파벳순 배열에 가깝다. 이용자들은 Medicine 내에 위치한 주제명을 모두 볼 수 있지만 항목의 상하관계, 연관관계의 논리적인 체계를 알아내기는 어렵다.

또한 정보자원 자체에 대해 이용자에게 설명해 줄 수 있는 기술방법에 있어서는 Yahoo!가 정보자원에 대한 주석 정도에 그치고 있는 반면에 Mr. Dui's Topic Finder 시스템에서는 인터넷 정보자원의 레코드를 다양한 필드로 등으로 구분하여 서지기술을 수행하고 있다.

3.3.4 비교분석 결과

이상과 같이 인터넷 기반 분류체계와 문헌분류체계를 인터넷 서비스상에서 비교한 결과 다음과 같은 사실을 확인하였다.

첫째, 분류항목의 집중도에 따른 분류대상 자원의 상이점이다. 인터넷기반 분류체계는 논리적이고 학술적인 분류에 따르기 보다는 이용자의 필요에 맞는 서비스가 우선적으로 개발되었음을 확인하였다. 둘째, 분류항목수의 차이이다. 인터넷기반 분류체계는 새로운 항목을 계속 추가하는 과정을 통해 증가되었음을 확인할 수 있었다. 셋째, 이용자 인터페이스의 차이이다. 문헌 분류체계는 안정된 분류를 따름으로써 이용자를 정확한 주제로 이끄는 방법을 구사하였고, 인터넷기반의 분류체계는 구체적인 주제명을 배열함으로써 이용자의 주제에 대한 직접적인 선택을 유도하였다.

이러한 상이점의 배경으로는 인터넷의 특성상 다양한 이용자 요구에 대처하기 위한 전략으로서 인터넷 기반 분류체계가 이루어졌다는 점을 들 수 있다. 계층성면에서 DDC가 체계적인 분류지침에 의하여 논리적으로 주제를 전개하는데 비해 Yahoo!와 같은 디렉토리 서비스는 여러 곳에 같은 주제를 배열하는 모습을 보이고 있으며, 변화능력 즉 유연성면에서도 DDC는 평균 약 6년마다 한번씩 개정되었음에 비하여 Yahoo!는 새로운 정보를 처리하는데 매우 역동적이며 끊임없이 새로운 주제가 추가되고 있어 변화에 매우 민감함을 알 수 있다.

이러한 비교결과는 문헌분류체계가 분류에 대한 체계적인 논리를 제공할 수 있는 점에 비하여 변화에 대한 대응이 늦어질 수 있음을 시사한다. 따라서 인터넷 환경에서 파생된 주제별 디렉토리 서비스들은 기존 분류체계와는 별도로 인터넷 정보서비스와 주제명 배열을 결합한, 이용자 서비스를 최우선 핵심역량으로 하는 시스템으로 파악하여야 한다.

4. 결론

본고에서는 인터넷 정보자원의 조직을 위한 분류개념의 적용을 인터넷기반 분류체계를 적용한 주제별 디렉토리화 기존 문헌 분류체계를

중심으로 조사하였다. 또 인터넷 상에서 이들 분류체계의 비교 평가를 통해 이상적인 분류체계의 모형을 제시하고자 하였다.

결론적으로 인터넷 정보자원의 효율적인 조직과 활용을 위한 이상적인 분류체계는 논리적이고 체계적인 주제분류와 함께 이용자의 다양한 정보요구와 이용행태의 분석을 통해 인터페이스 측면에서 다양한 접근점을 제공해 줄 수 있는 통합시스템임을 파악할 수 있었다. 주제의 전개과정은 계층적이고 논리적이어야 하지만 이용자의 검색 효율성을 제고하기 위해서는 변화에 민감한 유연성있는 체계의 적용이 필요하다. 따라서 분류체계의 설계에는 분류항목의 주제별 설계, 항목수의 설정, 전개방법, 확장가능성 등을 고려해야 한다. 특히 인터페이스 측면에서는 이용자가 쉽게 찾아갈 수 있는 접근점의 배열, 적절한 분류단계와 항목수, 공통항목과 순수주제의 구분을 통한 접근점을 적용할 필요가 있다.

또한 인터넷 정보자원의 효율적인 조직을 위해서는 이용자의 이용행태를 면밀히 분석 파악하여, 정보자원을 배열하는 연구가 우선적으로 수행되어야 할 것이다. 이 경우 이용자의 정보요구 분석결과를 토대로 원하는 주제에까지 직접적으로 접근할 수 있는 경로를 설계할 수 있을 것이다. 아직까지 국내외에서 이러한 인터넷 정보자원의 분류기준에 대한 연구는 초기 단계에 있으며, 최적의 분류모형을 설계하기 위한 지속적인 연구가 수행되어야 할 것이다.

참고문헌

- 김영보, 1997. "인터넷 탐색엔진의 분류체계에 관한 연구: 컴퓨터·인터넷분야를 중심으로," 성균관대학교 석사학위논문.
- 이명희, 1997. "네트워크 데이터베이스에서의 주제별 디렉토리화 키워드 탐색엔진의 탐색효율에 관한 탐색적 연구," 한국문헌정보학회지, 3(2) : 177-197.

- 정연경, 1996. "듀이십진분류표의 인쇄형과 전자형 비교 및 이용에 관한 연구" 정보관리학회지 13(2) : 97-119.
- 정영미, 1997, 지식구조론. 서울: 한국도서관협회.
- 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.
- Dahlberg, Ingtraut. "The Future of Classification in Libraries and Networks, a Theoretical Point of View," *Cataloging & Classification Quarterly*, 21(2) : 23-36.
- Floridi, L., 1996. "The Internet: Which Future for Organised Knowledge, Frankenstein or Pygmalion? Part 1 & Part 2," *The Library Journal*, 14(1) : 43-52.
- 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/>.
- Kent, R. E. and Neuss, C., 1994. "Creating a Web Analysis and Visualization Environment," <http://www.informatik.th-darmstadt.de/~neuss/www2/wwwf94.html>.
- 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>.