

지식분류체계의 변화와 동향

Trends of Knowledge Classification System

손소현(S.H. Son) 정보유통연구팀 연구원
홍기채(G.C. Hong) 정보유통연구팀 선임연구원
정현수(H.S. Jung) 정보유통연구팀 책임연구원, 팀장

최근 인터넷 산업이 부각되면서 분류에 대한 연구도 인터넷 분야의 다양한 응용에 적용하기 위해 새롭게 재기되고 있다. 본 고에서는 이러한 지식분류체계의 변화양상과 동향을 파악하기 위해 이론적 배경으로 가장 기본이 되는 도서관학 측면에서의 기존 분류체계의 특성을 살펴보았다. 이를 바탕으로 분류체계의 설계과정 기본원칙 및 새로운 적용분야에 대한 접근방식에 있어 분류체계에 대한 인식변화도 아울러 이루어져야 함을 살펴본다. 마지막으로 인터넷에 적용된 분류체계의 활용동향을 살펴보고자 한다.

I. 서론

분류(Classification)는 어떤 개념 또는 사건을 속성의 유사성에 근거하여 군집화하는 것으로, 논리적·지리적·연대·기원(유전적)·역사적(연대, 기원, 지리적 개념의 복합)·진화적(단순에서 복잡으로의 유사성 정도)·역동적·자모순·수학적(수치, 부호에 따른 것) 특성 등 다양한 속성에 따라 분류될 수 있다 [1-2].

이러한 분류는 장서 및 데이터베이스 관리, 시소러스 구축 및 유지 등의 정보관리기능의 토대가 될 뿐만 아니라 정보접근도구(Accessing Tool), 브라우징(Browsing), 항해(Navigating), 확인(Validation), 전환(Transformation) 및 매핑(Mapping), 평가(Evaluation), 필터링(Filtering) 등 정보검색 결과를 강화하는 데 활용되기도 한다.

초창기 분류연구는 분류표와 주제명표목표 등의 정보자료의 서가배열목적에 한정되어 있었고, 그 방법도 수작업으로 분류하는 방식을 택했다. 그러나,

최근 인터넷이라는 새로운 적용분야가 부각되면서 분류에 대한 연구가 인터넷 분야의 다양한 응용에 분류를 적용하기 위해 새롭게 재기되고 있다. 또한, 분류방법도 자동화시켜 분류하는 방식을 취하고 있다.

본 고에서는 이러한 지식분류체계의 변화양상과 동향을 파악하기 위하여 이론적 배경으로 가장 기본이 되는 도서관학 측면에서의 전통적인 분류체계의 특성과 새로운 분류체계의 모색 방안에 대해 분석해 보았다. 또한, 도서관학 측면에서 뿐만 아니라, 인터넷에 활용되는 분류체계에 모두 적용되는 지식분류체계의 설계과정 기본원칙을 살펴보았다.

기존의 지식분류체계에 비해 현재 새로운 분류대상으로 부각되고 있는 인터넷 자원들은 비정형적이며, 수시로 유동적으로 변화하는 특성을 가지고 있다. 새로운 분류대상에 대한 지식분류체계의 새로운 접근방식으로, 정보자체의 환경변화를 유연성있게 수용하기 위해 지식분류체계에 대한 인식변화 양상도 아울러 변화되어야 함을 살펴보고 인터넷 정보자원에 대한 분류체계 활용동향을 살펴보고자 한다.

II. 분류의 이론적 배경

1. 도서관학 측면에서의 기존 분류체계 특성

분류는 사용하려는 목적에 따라 학문분류와 문헌분류로 나눌 수 있다. 학문분류는 학문 자체의 분류를 목적으로 혹은 사물이나 개념 상호간의 관계를 발견하는 수단으로 사용하는 분류이며 문헌분류는 문헌의 효과적인 이용을 위해 체계적으로 정보를 재배치하는 도구로서 활용하려는 분류이다. 현재까지 도서관학 측면에서 보는 기존 분류체계는 대부분 문헌분류에 초점을 맞추어 구성되어 있다.

가. 대표적인 문헌분류체계

문헌분류는 구조원리상 일반적으로 열거식, 계층식, 분석합성식 등 세 가지 분류방법으로 구성된다. 열거식은 연속적으로 개념범주가 좁아지면서 특정 주제로 분할하는 방법으로 분할된 모든 주제에 대해 표목을 할당하고 그것을 알파벳순으로 열거한다. 계층식은 분류될 주제의 기본 조직에 근거해 주제를 류, 과, 속의 계층구조로 나누어 표현한다. 분석합성식은 용어를 개별개념으로 분할하고, 분할된 개별개념에 특정 규칙을 제시하여 합성주제를 만들고 이를 새로운 표목으로 구성하여 나타낸다.

대표적인 문헌분류체계는 DDC(Dewey Decimal Classification), UDC(Universal Decimal Classification), LCC(Library of Congress Classification)가 있다. 기존의 분류체계 중 가장 오래된 DDC는 1876년 소규모 교양과목장서를 정리하기 위해 듀이(Dewey)가 만든 것으로, 해리스(Harris)의 분류체계(베이컨(Bacon)의 지식이론을 바탕으로 설계)를 기초로 하여 작성되었으며 전반적으로 이용효율을 강조한 분류체계이다.

UDC는 1895년에 출판되었으며, 전세계 모든 인쇄기록 지식을 분류하기 위한 국제적인 색인 작성을 목적으로 작성되었다. DDC 보다 더욱 지식을 세분화 하였다는 특징을 가지며, 표준화된 하위구분과 기호를 사용하였다.

1904년 만들어진 LCC는 진화의 순서원칙을 강조한 C.A Cutter의 EC(Expensive Classification) 분류체계의 주류배열을 참조하여 21개의 배열로 표현하였다. LCC의 가장 큰 장점은 주제관련 하위구분을 나타내는 방법에 있어 문자로 시작하는 폐쇄된 연속번호를 이용하여 개념을 고정시키고 결합하는 등 다양한 하위구분 방법을 사용할 수 있다는 것이다. 하나의 학문 구성요소를 하위 구분하는데 대체로 공식적이고 다양한 방법을 제시하였으며, 다른 분류체계에 비해 상대적으로 짧은 분류기호로도 충분히 분류를 표현할 수 있다. 반면, 이러한 분류기호들이 조기성을 가질 수 없으며 미리 결합된 개념들은 차후 분석적으로 탐색될 수 없기 때문에 이러한 특성이 일부 단점으로 제기되기도 하였다.

대표적인 문헌분류체계인 DDC와 LCC 분류체계의 또 다른 몇가지 특징들을 요약하면 아래와 같다. 첫째, 현재 사용되고 있는 문헌분류체계가 거의 모든 주제를 포괄한다. 둘째, 대부분의 분류체계는 학문에 기초를 두고 있어 항목 전개상의 논리성과 주제 선정이 명확하다. 셋째, 항목 전개의 논리성과 높은 검색 효율을 입증할 수 있도록 주제의 배열이 계층구조를 가지고 있다. 넷째, 항목을 조절하고자 할 경우 미리 기술된 분류표의 하위항목이나 보조표 등을 이용하여 쉽게 확장할 수 있다.

나. 대표적인 분류원칙

전통적인 분류는 명확한 분류체계원칙을 확립하고 그 원칙에 따라 수작업으로 분류하는 방식을 택했다. 따라서, 분류원칙은 분류체계 구축에 있어 중요한 역할을 담당한다. 1910년 발표된 블리스 이론은 전체 32개의 분류원칙을 제시하였으며, 그중 가장 비중이 높은 세 가지 원칙을 살펴보면 아래와 같다. 첫째, '교육적이고 과학적인 일치'의 사고표현으로서 교육자, 과학자, 철학자에 따라 인지하고 있는 지식을 구조화한다. 둘째, '관련된 주제의 배열'로서 주제는 파생된 류들을 묶는 것이다. 셋째, '전문성의 등급'으로서 가장 일반적인 것에서 특수한 것까지 동등한 주제를 순서대로 조직하는 것이다[3].

1933년 출판된 랑가나단의 CC(Colon Classification)는 ‘증가하는 구체성 원칙’과 ‘감소하는 자연성 및 증가하는 내용의 인공성 원칙’에 따라 분류를 배열한다. 다섯 가지 기초적인 패킷(P, M, E, S, T)으로 나누어 각 패킷이 각 주제와 대응되도록 하며, 대응된 각 패킷은 개별적인 열거순서 또는 패킷공식을 가지고 있다. 특히 랑가나단은 분류요소, 주제종류, 분리와 범주, 패킷방식을 이용한 특수한 종류의 개념조합 관계를 설정하여 패킷방식이 예상 가능한 순서로 분류표의 모든 항목을 결합할 수 있는 가능성을 보여주었다. CC 원리는 자동화된 분류시스템 구축을 위한 이론적 지침이 되었다[4].

2. 새로운 분류체계의 모색

가. 패킷분류

보통 전통적인 분류체계는 기본적으로 열거식 분류이지만, 새로운 분류체계에서는 많은 부분에서 패킷요소를 포함하고 있다. 랑가나단의 CC에서 처음 시작된 패킷분류는 분석합성식 분류로서 여기서 분석은 상위요소가 패킷분석을 통해 정의된 단일개념으로 분해하는 것이고, 합성은 그러한 개념을 주제 문자열로 재결합하여 새로운 요소가 만들어지는 것이다. CC는 다섯 가지의 기본적인 패킷범주와 쿨론을 포함하며, 합성된 새로운 분류기호 내에 포함된 단일개념의 패킷과 기호체계가 패킷지시자로서의 역할을 한다. DDC와 LCC 등의 전통적인 분류체계가 모든 미디어의 모든 주제를 정확하게 기술하는 데 충분한 기술적 바탕을 제공하지 못해 개별도서관이나 특수도서관의 요구를 충분히 수용할 수 없었던 점과 용어의 충분한 조합을 제공하지 않고, 종종 자료를 찾기 어렵다는 이유 때문에 현재 패킷분류가 많은 관심의 대상이 되고 있다.

나. 패킷분류와 시소러스와의 통합

현재 사용되고 있는 대부분의 분류표는 부분적으로나 전반적으로 패킷에 기초해 구축되어 있는데, 최근에는 패킷분석 과정이 시소러스를 구축하는 데

사용될 수 있어 패킷분류와 시소러스와의 통합문제도 중요한 연구대상이 되고 있다.

패킷분류와 시소러스의 결합에 대한 연구는 1969년 에이치슨(J. Aitchison)에 의해 개발된 전기공학 분야의 “Thesaurofacet”과 1994년 미국의 Getty Art History Information Program의 지원 하에 만들어진 “AAT(Art and Architecture Thesaurus)”가 대표적이다. 이 두 연구에서는 기본적으로 모든 분류 부문은 종속관계와 같은 완전한 계층으로 묘사되며, 시소러스 부문은 상위어(BT), 하위어(NT), 관련어(RT)와 같은 계층 부문들간의 용어관계를 보여주고 있다. 전통적으로, 패킷분류는 전조합에서만 사용될 수 있고, 시소러스는 후조합에서만 사용되지만 본 연구들에서는 패킷분류표가 자료를 목록하는 주제표목을 만드는 데 사용되었으며, 시소러스는 데이터베이스로부터 후조합방식으로 자료를 검색하는데 사용될 수 있게 함으로써 시소러스와의 효율적인 통합방법을 제시하였다.

III. 새로운 정보환경과 분류

1. 분류체계의 설계과정원칙

앞서 제시된 도서관학 측면에서의 전통적인 분류 방법 및 새로운 분류방법 뿐만 아니라 인터넷에 활용되는 분류체계에 모두 적용되는 공통적인 목적 및 설계과정원칙이 있다. Marcello와 Newton은 분류의 가장 중요한 목적으로 정보 검색자에게 최대한 유용한 주제어 리스트를 제공하는 것으로 본다[5]. 다시 말해 분류는 수집된 자료를 이용자에게 효과적으로 제공하는 것이 목적이 되어야 하며, 분류체계를 설계할 때 설계자는 이용자가 가장 쉽고 정확하게 자료에 접근할 수 있도록 하는데 중점을 두고 설계해야 한다.

분류체계의 설계과정은 ① 주제선정 및 주제구조 방법으로 나타내는 개념적 단계, ② 주제를 개별적인 단어로 나타내는 용어학적 단계, ③ 모든 주제에 다른 코드를 부여하는 기호화 단계로 구성된다. 이러한 각 설계단계에는 다음과 같은 원리와 규범이

요구된다[3].

가. 지식의 개념구조 단계

1) 특성구분

분류체계에서의 지식구분은 일상생활에서 사물을 범주화하는 것과는 달리 다음과 같은 방법으로 이루어져야 한다. 첫째, 분류간에 영역이 분명해야 한다. 둘째, 분류체계의 범주는 한번 선정되면 영구적이어야 한다. 셋째, 분류체계에서 지식을 구분하기 위한 특성은 모든 사람들이 입증할 수 있고 이용 가능해야 한다. 넷째, 특성은 분류목적에 적절해야 한다.

2) 특성순서

분류체계에서 지식을 구분하는 것은 단지 하나의 특성만으로는 충분하지 않으며 여러 가지 특성을 사용하여 지속적인 하위범주화가 필요하므로 서로 배타적이며, 동시에 하나의 특성이 더 작은 범주로 구분될 수 있도록 적용특성을 명확히 순서화 하여야 한다.

3) 동위류의 조직

동위류는 어떤 수준에서든지 지식구조 내의 모든 개념을 수용할 수 있어야 한다.

4) 하위류의 규범

하위류로 갈수록 관점은 좀더 좁아지고 깊이는 좀더 깊어져야 한다.

나. 용어학적 단계

- 1) 가장 최신의 정확한 주제명 사용
- 2) 모든 주제에 대해 중립적인 용어 선정
- 3) 기술적인 용어를 배제한 단순화된 용어 선정

다. 기호화 단계

분류체계는 이용될 용어의 선정보다 개념적 조직과 기호가 중요하다. 기호는 분류체계에서 개념과 주제에 부여되는 코드체계이며, 기호에는 다음과 같은 원칙이 있다.

- 1) 계층성: 이용자에게 친숙하고, 간략하며, 개념적인 계층을 표현할 수 있는 기호 사용
- 2) 간결성: 최대한 중복을 피하여 간결하게 표현

2. 정보환경의 변화와 분류에 대한 인식

지식은 지속적으로 분열되면서 동시에 통합되는 역동적인 연속체로서, 도서관에서는 대체로 문헌분류체계를 이용하여 지식을 분류하고 있다. 그러나 학문의 역동적인 성향과 함께 인터넷이라는 새로운 정보형태가 개발되고 보급됨에 따라 학문의 갱신 주기도 점차 짧아지고 있다. 이에 따라 전통적인 분류와 주제분석기법을 이용한 분류방법에 여러 가지 문제점이 대두되고 있다.

전통적인 지식분류는 대부분 학문영역에 근거하고 있으나, 최근 새로운 영역의 학문이 생성되고 기존 학문과 통합되는 경향의 증가로 인하여 대체로 종래의 엄격한 기준이 적용되지 않는 실정이다. 이러한 유연성을 지닌 학문영역을 수용하기 위해 지식체계는 지식을 하위요소로 세분하는 분류화와 동시에 지식의 점진적인 통합과정으로 새로운 학문영역을 만드는 개념적 통합화를 표현할 수 있어야 한다[3]. 이러한 다면적 관계를 하나의 체계 하에 표현하는 기법에 대한 연구가 활발히 이루어져야 할 것이다.

또한, 현재 사용되고 있는 분류체계는 다학문적(interdisciplinary)인 문헌과 두 가지 이상의 주제를 동시에 가진 문헌을 각 주제분야에 중복 분류할 수 없는 단일 분류체계를 가지는 것이 대부분이므로, 정보시스템이 지속적으로 규모가 커지고 확고해짐에 따라 기존의 분류와 주제분석방법이 새로운 요구를 수용할 수 있도록 재구성되어야 할 것이다.

또한, 정보처리기술의 발전과 인터넷 이용 확산과 더불어 인터넷을 통해 얻을 수 있는 정보의 양은 방대해지고 있다. 아울러 정보자원을 체계적으로 조직하여 필요로 하는 정보만을 쉽고 빠르게 찾을 수 있도록 도와주는 다양한 검색도구들이 등장하고 있다. 이에 발맞추어 효율적인 검색 인터페이스 및 내부처리방안으로 분류체계를 활용하는 것에 대한 관심이 증가하고 있다. 반면, 기존 분류가 주로 인쇄자

료와 서지자료를 조직하기 위해 설계되어 있어, 오늘날의 인터넷을 포함한 전자환경에서의 원격자원과 정보소스의 위치에 대한 정보조직화를 반영하는 등 일부 재조정이 요구된다.

IV. 인터넷 정보자원의 분류체계

1990년도 초반 해도 국내에서 정보검색시스템을 도입한 곳은 한국산업 정보 연구원, 외무부, 국회 도서관, 대우 고등 기술원, ADD, 문화 방송국 등으로 주로 기술 정보나 사규 관리, 도서 관리 등의 업무에 적용되어 왔다. 그러나 현재는 인터넷과 인트라넷의 도입 및 그에 따른 영향으로 각 회사나 공공기관에서 이러한 정보검색시스템을 웹으로 제공하고 있으며, 인터넷 산업이 크게 확장되고 엄청난 양의 지식 정보가 쏟아짐에 따라 인터넷 이용성향은 좀더 편리하고 개인 중심적인 서비스로의 경향을 띄고 있다. 이제는 얼마나 많은 데이터를 보유하는가 보다 이용자에게 유용하고 정확한 정보를 효과적으로 서비스하는 기술관점으로 초점이 바뀌고 있다. 다양한 정보를 수집하여 사용자의 요구에 적합한 정확한 정보를 신속하게 제공하는 것이 현재의 정보검색시스템이 내세우고 있는 목표이다. 새롭게 대두되고 있는 분류 검색기술은 정보 이용자가 새로운 지식에 처음 접하는 경우에도 관련어 및 유사어 인식이 가능하며 계층적인 구조를 쉽게 파악할 수 있어 초기개념 인식부담이 감소한다는 이점이 있다.

1. 기존의 문헌분류체계의 적용

앞서 기술한 DDC, UDC, LCC 등의 대표적인 문헌분류체계가 도서관장서의 분류를 위해 널리 사용되고 있음에도 불구하고 웹 문서 분류체계에 적용되지 못하는 인터넷에 등재된 문서들의 주제가 대부분 새로운 주제분야를 다루는 것이 많아 용어가 없거나 주요한 개념은 대부분 류·강의 수준에서 배열되고 고정적인 형태를 취하고 있다. 특히 새롭게 생성되는 개념과 주제에 대해서는 세목과 그 하위의 분류

기호를 부여하여 확장하고 있긴 하지만 새로운 지식이나 주제가 실시간으로 생성되는 인터넷의 특성을 기존 분류체계로 수용하는 데 현실적 어려움이 많다. 따라서 기존의 문헌분류체계를 웹 문서 분류체계로 그대로 적용하기 보다 인터넷상의 정형화, 체계화되지 않은 정보의 특성상 효율성이 높은 정보검색시스템 구축이 이루어 질 수 있도록 체계적이고 논리적인 기존 분류체계 특성을 일부 수용할 수 있는 방안으로 활용되어야 할 것이다.

Svenonius(1983)는 온라인 환경에서 분류체계의 적용이 필요함을 주장하고, 분류체계를 이용하는 것이 검색의 재현률과 정확률을 향상시키며 탐색에 대한 문맥 및 의미적 브라우징 기능을 제공할 뿐만 아니라, 상이한 언어간의 변환을 위한 메커니즘 및 용어간의 상관관계를 계층적으로 보여주는 데 도움이 된다고 제시하였다[6].

그 후, Markey와 Demeyer(1996)는 DDC 온라인 프로젝트에서 이용자의 주제접근, 브라우징, 배열을 개선시키기 위한 분류체계의 구현방안을 제시하였다[7]. 온라인 환경에서의 검색도구로 분류의 유용성에 대한 검토가 고려된 것은 최근의 인터넷 이용의 증가로 인해 OCLC를 중심으로 인터넷 정보자원에 대한 브라우징 구조 및 검색능력을 향상시키는 도구로 분류가 인식되면서부터 시작되었다.

Vizine-Goetz(1996)는 DDC와 LCC의 분류체계를 탐색엔진 야후(Yahoo)의 분류체계와 비교, 분석하여 각 항목들의 용어와 구성에 포함된 수를 조사하여 항목간의 균형성과 적절성에 대한 문제를 밝혀냈다. 그 후, Vizine-Goetz(1997)는 OCLC에서 이러한 분류체계를 적용한 사례를 레코드 형식과 구축방법에 대한 소개와 함께 제시하였다[8].

Koch(1997)는 분류체계에 의한 인터넷 정보자원 조직의 장점을 다음과 같이 요약하였다[9].

- ① 브라우징(Browsing): 분류된 주제들을 쉽게 브라우징 할 수 있다. 특히 해당 주제의 내용, 구조, 용어 등에 익숙하지 않은 이용자들에게 큰 도움이 된다.
- ② 탐색의 확대와 축소(Broadening and Narrowing

- Searches): 분류표는 계층구조로 되어 있으므로 필요시 탐색의 확대와 축소가 용이하다.
- ③ 문맥(Context)의 제공: 분류표는 검색에 이용될 용어에 문맥을 제공한다. 예를 들어 동형이어의 문제를 부분적으로 해결해 준다.
 - ④ 다국어 접근의 가능성(Potential to Permit Multilingual Access): 분류표에서는 숫자기호만을 사용하므로 다국어 자료 접근이 용이하다.
 - ⑤ 분할 및 조작(Partitioning and Manipulation): 분류표의 큰 주제항목은 필요시 논리적으로 작은 주제항목으로 세분되거나 조작될 수 있다.
 - ⑥ 주지성(Potential to be well-known): 많은 도서관 이용자들은 전통적인 도서관분류체계에 익숙해져 있다. 이들은 자기의 관심 주제분야에 대한 분류표 분류전개에 익숙해져 있을 가능성이 높다.

이상과 같이, 기존의 문헌분류체계를 이용하여 인터넷 정보자원의 정보검색시스템으로 활용하는 주요 목적이 검색의 정확도를 향상시키고, 부가적인 서비스형태로 제공하는 것으로 바뀌고 있다. 검색의 편의성 제공을 향상시키기 위해 주제분류 위주의 브라우징 구조를 제공하며, 검색시스템의 검색 처리 속도 향상 및 정확도 향상을 위해 검색결과를 주제별로 분리할 수 있도록 연산기능을 제공하는 등 다양한 접근 방법을 사용하고 있다.

2. 인터넷 검색엔진에 응용한 분류체계

가. 디렉토리 서비스 분류체계

현재 인터넷상에 제공되는 검색엔진을 구분하는 공통된 기준은 없으나 대체적으로 키워드 검색엔진, 디렉토리 검색엔진, 메타 검색엔진으로 구분하는 것이 일반적이다. 키워드 검색엔진은 주제어 검색엔진과 동일한 의미로, 하나의 데이터베이스에 모든 URL과 특정 키워드를 입력함으로써 원하는 정보를 찾는 검색엔진이다. 디렉토리 검색엔진은 주제별 검색엔진과 동일한 의미이며, 특정 주제별로 각 페이지들을 분류하여 정리해 놓은 것으로 웹사이트의 주제별 목록을 제공하는 검색엔진이다. 메타 검색엔진은 순

차적 검색엔진(Sequential Search Engine)이라고도 하며, 많은 검색엔진들을 모아 검색하기 편하게 만든 도구로써, 여러 검색엔진에 대한 단일화된 인터페이스를 제공하여 검색결과를 종합해 주는 기능을 제공하므로 별도의 데이터베이스를 가지고 있지는 않다.

여러 유형의 검색엔진 가운데 직접적으로 분류체계를 이용하는 디렉토리 검색엔진은 웹사이트를 일정한 체계에 따라 계층적으로 분류하여 목록을 제공하며, 배열된 주제에 따라 이용자가 찾고자 하는 정보의 개념을 논리적인 전개와 구체적인 주제항목으로 제시해 주는 이용자 편의 지향적인 서비스이다. 최근 수많은 사이트들이 인터넷 정보자원에 대한 효율적인 접근방법으로 디렉토리 서비스의 유용성을 인정하여 이를 도입하고 있다. 대표적인 종합 포털 사이트인 야후, 알타 포탈 등의 검색엔진들은 자체 구축한 분류체계를 이용하여 주제별 디렉토리 서비스를 제공하고 있으며 사이트에 대한 간단한 설명에서부터 관련링크, 논평, 등급 평가기능을 추가 제공하거나, 디렉토리 내의 키워드 검색기능을 제공하기도 한다.

나. 시소러스

정보검색시스템에서 분류체계를 활용하는 방법으로 초보자에게 검색 편의성을 제공하는 디렉토리 서비스 외에 검색의 정확도를 높이기 위한 도구로 활용되는 분류체계가 있다. 이용자가 검색을 하면 검색 질의어로 던진 용어만을 이용하여 검색하는 것이 아니라, 해당 용어와 비슷한 의미를 가진 다른 용어들이 있을 때에 관련된 용어를 함께 보여주거나, 해당 용어가 언어학상 여러 가지 의미를 가질 경우 각각의 의미와 관련된 용어들을 각각 함께 보여줌으로써 궁극적으로 얻고자 하는 결과에 근접한 문서를 얻을 수 있다.

이를 위해서는 색인단계에서부터 고려가 이루어지는 방법과 검색단계에서 고려가 이루어지는 방법이 있다. 색인단계에서 고려가 이루어지는 방법은 일명 전조합방식으로 색인시 의미가 유사하거나 동일한 용어를 서로 연관시켜 각기 다른 용어로 색인

되는 것을 미리 방지한다. 검색단계에서 고려하는 방법은 후조합방식으로 색인시에는 각 용어별로 색인하고 검색시 용어간의 관계를 조합하여 정보를 검색할 수 있도록 한다.

용어간의 관계를 정의하는 분류체계를 시소러스라고 하며, 시소러스는 대표어와 관련용어 리스트로 이루어진다. 관련용어의 종류로는 용어(Term), 주제분야(SUBJ), 우선어(USE), 비우선어(UF), 상위어(BT), 하위어(NT), 관련어(RT) 등이 있을 수 있다. 시소러스가 포함한 관련용어의 종류에 따라 단순히 동의어를 통제하는 단계에서부터 계층관계 및 주제분야 정보를 통제하는 단계까지 다양하게 활용할 수 있다.

초기의 시소러스는 수작업 서지 데이터베이스에서 사용되었지만, 현재는 거의 모든 검색시스템에서 전문검색을 지원하고 있고 방대한 인터넷 정보를 다루고 있기 때문에 자동화가 필수불가결하게 되었다. 또한, 형태소분석을 근간으로 하는 자연어처리기능을 대부분 내장하고 있어 전문에서 각 용어추출이 가능하게 되었고 이들에 대한 분석을 통해 기술적으로 시소러스 구축을 자동화하는 것이 가능하게 되었다.

시소러스는 구축뿐 아니라 갱신에도 자동화가 필요하다. 인터넷이 정체상태가 아닌 이상 색인과 시소러스 갱신은 지속적으로 이루어져야 한다. 이런 지속적인 갱신이 이루어지는 시소러스를 기존의 정적인 시소러스에 반하는 의미로 동적 시소러스(dynamic thesaurus)라고 부른다. 갱신을 위해 자동분류를 이용하여 시소러스를 편집하는 것은 계산량이 많아 시간이 많이 소요된다. 동적 시소러스에 관한 연구는 용어간의 관계를 어디에서부터 추출하는가에 따라서 다양하다. 리스포터(L.K. Rees-Potter)는 인용문맥 분석을 이용하여 시소러스를 갱신하고 유지하는 체계를 연구하였다. 권췌 등(U. Güntzer *et al.*)은 개념들간의 관계가 문헌에 나타나기 전에 정보검색시스템의 이용자들이 먼저 알고 사용한다는 가정에 근거하여 실제 검색에서 이용자에 의해 조합되는 용어들을 분석하여 이용자의 전문 지식을 추출하고, 추출된 지식을 시소러스에 결합시킴으로써 용어간 관계를 갱

신하였다. 키모토(H. Kimoto)와 이와테라(T. Iwadera)는 이용자의 특별한 관심사를 반영하는 동적 시소러스를 구축하였다. 이용자가 자신의 관심주제에 적합한 문헌을 선택하게 한 다음, 이 표본 적합문헌으로부터 용어정보를 자동으로 추출하여 개인별로 동적 시소러스를 구축하였다[10].

V. 결론

지금까지 지식분류체계의 변화와 동향을 파악하기 위하여 기존 분류체계의 특성과 새로운 분류체계의 모색에 관해서 조사하고, 분류체계의 설계과정 원칙과 정보환경의 변화에 따른 분류에 대한 인식의 변화 양상을 검토한 후, 인터넷 정보자원의 분류체계 활용동향에 대해서 살펴보았다.

최근 기존의 인터넷 검색엔진 분류체계에 대한 이용자의 요구사항을 조사한 결과, 이용자 특성을 제대로 파악하고 있지 않으며, 주제접근을 위한 주제명이 표준화되어 있지 않고, 주제어 수도 부족하다고 평가되었다. 또한 주제어 선택을 위한 다양한 인터페이스를 갖춘 브라우징 기능이 요구되며, 주제어 선정시 문헌분류체계의 참조가 필요함을 지적하고 있다. 따라서, 분류를 활용한 인터넷 정보자원의 조직화는 주제접근 도구로서 다음의 특징을 가지고 있어야 할 것이다. 첫째, 정확한 주제접근과 검색용어의 문맥적인 의미를 제공하여 동일용어의 경우 관련주제를 제시함으로써 의미의 혼동을 최대한 감소시켜야 할 것이다. 둘째, 유동적인 지식분야를 수용할 수 있도록 지속적으로 갱신되어야 한다. 셋째, 이용자에게 편리한 인터페이스로 브라우징 기능을 제공하여야 할 것이다.

참고 문헌

- [1] Amand, Maple, "Faceted Access: A Review of the Literature," http://www.misic.indiana.edu/tech_s/mla/facacc.rev
- [2] 정영미, "지식구조론," 한국도서관협회, 1997.
- [3] Hemalata. Iyer, "Classificatory Structures: Concepts,

- Relations and Representation,” Frankfurt/Main: IND EKS Verlag, 1995.
- [4] Ingetraut. Dahlberg, “The Future of Classification in Libraries and Networks, a Theoretical Point of View,” *Cataloging & Classification Quarterly*, Vol. 21, No. 2, 1995, pp. 23 – 36.
- [5] R. Marcello and R. Newton, “A New Manual of Classification,” New york: Gower Publishing, 1994.
- [6] E. Svenonius, “Use of Classification in Online Retrieval,” *Library Resources & Technical Services*, Vol. 27, No. 1, 1983, pp. 76 – 81.
- [7] D. Vazine-Goetz, “Using Library Classification Schemes for Internet Resources,” 1996, <http://www.oclc.org/oclc/man/colloq/v-g.htm>
- [8] D. Vazine-Goetz, “OCLC Investigates Using Classification Tools to Organize Internet Data,” 1997, <http://www.oclc.org/oclc/news/n226/research.htm>
- [9] R. Denenberg, “Structuring and Indexing the Internet,” 1996, <http://lcweb.loc.gov/z3950/agency/italy.htm>
- [10] 이재윤, “동적시소러스의 구축에 관한 실험적 연구,” <http://dewey.yonsei.ac.kr/memexlee/doc/thes.htm>