

인터넷 검색엔진의 디렉토리 구성에 관한 연구

A Study on the Organizing Directory for Internet Directory Search Engines

신동민(Dong-Min Shin)*

초 록

본 연구는 디렉토리 검색엔진의 주제구성에 대한 문제점을 분석하고, 효율적인 디렉토리 구성과 운영을 위한 지침을 제시함으로써 인터넷 공간에서의 정보검색에 효과성을 높일 수 있도록 하였다. 연구방법으로는 연구대상인 디렉토리 검색엔진의 최상위 주제와 하위 주제를 조사하여 검색엔진들의 일반적 특성과 주제구성의 계층성, 논리성 등을 분석하였고, 이론적 배경을 위해 문헌적 연구를 병행하였다. 연구 결과로 디렉토리 구성, 주제 선정, 인터페이스, 유저·관리 등을 위한 10가지의 지침을 제시하였다.

ABSTRACT

The purpose of this study is to suggest the guidelines for organizing and maintaining their subject directory search engines which serve effective search results of web documents.

The methods of this study are to review and analyze some directory search engines for finding problems of them and review literatures concerned with classification theory and previous studies. As results, this study suggests the guidelines for preparing the systematic subject directory scheme that is adapted to the internet directory search engines related to general and/or special subject fields through using the above guidelines.

키워드: 디렉토리 검색엔진, 검색엔진, 인터넷, 주제명

internet, directory search engine

* 한남대학교 문과대학 문헌정보학과 부교수(domishin@mail.hannam.ac.kr)

- 논문 접수일: 2001년 5월 15일
- 게재 확정일: 2001년 6월 7일

1 서 론

우리는 지금 무한경쟁시대에서 살고 있다. 경제, 사회, 문화, 교육, 심지어 종교에 이르기 까지 남보다 잘하지 않으면 도태되거나 뒤지게 되어 즐거운 삶을 영위할 수 없게 되었다.

이런 경쟁사회의 최 첨병은 역시 정보가 재화로서 인정되고 있는 정보화사회에서의 정보 서비스 산업일 것이다. 그 중에서도 정보의 보고, 바다라 할 수 있는 인터넷 공간에서 이용자가 필요한 정보를 찾을 수 있도록 도와주는 검색엔진이야말로 이용자의 관심을 끌지 못하면 아무리 훌륭한 포털 기능을 갖추고 있다고 하여도 경쟁에서 뒤지게 되며, 결과적으로 사라지게 될 것이다.¹⁾ 따라서 향후 인터넷상의 웹 문서를 찾을 수 있도록 도와주는 검색엔진의 부침(浮沈)이 많을 것으로 예상되며, 각각의 검색엔진들은 그들 나름대로의 생존전략을 수립해야 할 것이다. 특히 모든 주제를 다루는 일반주제 디렉토리 검색엔진은 인터넷 공간의 정보 증가 속도로 볼 때, 점차적으로 검색엔진의 2차적인 정보를 제공하게 될 것으로 예상되며, 이미 주제전문 검색엔진의 수가 상당한 수준으로 증가하고 있음을 볼 수 있다.

실제로 정보의 바다 또는 정보의 보고라 일컬어지는 인터넷상의 웹 문서로부터 이용자가 필요한 정보를 정확히 찾기란 참으로 어렵다. 그 이유 중 가장 비중이 높은 것은 역시 웹 문서 작성자, 즉 정보 생산자들이 일정한 기준 없이 임의로 정보를 생산하여 유통시키고 있으므로, 웹 문서를 탐색하여 이용할 수 있는 수단인 웹 검색 엔진들도 필요정보의 탐색에 정확성을 유지할 수 없는 것이다.

특히 웹 문서들을 일정한 방법으로 범주화하여 구분함으로써 탐색자들에게 정보접근을 가능하게 하는 디렉토리 검색엔진들은 류 개념인 최상위주제로부터 종 및 속 개념의 하위계층에 이르기까지 독자적인 분류체계를 갖고 있어서 이용자들은 주제접근에 어려움을 겪고 있다.

따라서 본 연구는 이용자의 입장에서 현재 이용자들이 정보탐색에 이용하고 있는 디렉토리 검색엔진들의 주제분류체계를 조사하여 분석함으로서 합리적이고 체계적인 디렉토리 구성을 위한 원칙을 제시하여 웹 문서 분류체계 작성 지침으로 활용할 수 있도록 힘을 목적으로 하고 있다.

이러한 연구는 인터넷을 통한 정보접근에 익숙하지 않은 이용자라도 이미 익숙한 문헌분류 체계라든지 사회통념상의 계층적 주제구조 등이 합리적이고 체계적이며 표준화된 방법으로 조정됨으로써 탐색도구들을 쉽게 이용할 수 있도록 하는데 도움을 줄 것이다.

본 연구의 연구방법으로는 지식 및 문헌분류 체계를 검토하여 주제분류의 원리와 계층구조를 밝히고, 현행 국내 디렉토리 검색엔진들의 주제분류체계를 조사하여 분석 하므로서 개선 방안을 도출하고자 한다.

연구범위로는 디렉토리 검색엔진에 적용될 수 있는 주제어의 범위와 계층관계, 비 주제분야와 관련된 조건 제한 검색방법, 검색엔진에서 제공되어야 할 주제접근 용이성 등에 대한 지침으로 활용할 수 있는 내용을 제시하는 것 까지로 한다. 따라서 류, 종 및 속 개념에 속하

1) 실제 본 연구 진행 중 “Aza”는 2001년 5월 10일자로 서비스를 중단하고 검색엔진의 수명을 다하였다.(서비스 종료의 이유는 “지속적인 유지가 어려워” 이었다.)

는 주제어 선정 및 주제어간 상호관계 등을 구체적으로 제시하는 것은 범위에서 제외한다.

조사대상 검색엔진으로는 국내에서 비교적 자주 이용되는 것으로 소개되고 있는 것 중 야후코리아, 심마니, 한미르(정보탐정), 네이버, 까치네, Search-korea, Lycos-korea, Zip 등 8개로²⁾ 정하였다.

2 선행연구 개관

인터넷상의 온라인 정보검색은 정형화, 체계화되지 않은 정보의 특성상 효율성 높은 검색시스템이 필요하다. 이러한 필요성에 의해 현재 온라인 검색시스템에 체계적이고 논리적인 분류체계 적용이 검색효율을 높일 것이라는 추정 하에 많은 연구가 진행되고 있음을 볼 수 있다.

먼저 1983년에 온라인 검색에서 분류체계 적용이 검색 효율성(신속성, 정확성, 경제성)을 증진시키는데 유용하다는 연구가 Svenonius(1983)에 의해 이루어졌다. 이어서 Markey와 Demeyer(1986)도 DDC 온라인 프로젝트에서 최종이용자의 주제접근, 브라우징, 배열에 있어서 도서관 분류체계의 탁월한 유용성을 역설한 바 있다. 1995년에 수행된 연구에서 Dahlberg(1995)는 네트워크 환경에서 LC와 DDC를 적용해서 분류이론 적용에 대한 적합성 여부를 분석하였는데, 각 주제의 분석과 추적 및 계층구분에 유용함을 밝혀냈다.

1996년에는 Vizine-Goetz(1996)가 인터넷 정보에 대한 주제분류체계에 대해서 DDC 및 LC의 분류체계와 YAHOO를 비교·분석한 결과 주제어에 링크된 문서수의 항목간 균형이

있음을 밝혀냈다.

또한 Traugott(1997)는 인터넷 정보자원의 조직에 문헌 분류체계를 적용할 때의 장점에 대해서 주제세분과 조작, DDC 분류체계에 대한 인지도 등의 주지성의 장점이 있다고 하였다.

국내에서는 1997년 김영보(1997)가 문헌분류이론을 적용해서 인터넷 탐색엔진에 적용할 분류기준을 정립하고 정립된 분류기준으로 기존의 인터넷 검색엔진들의 컴퓨터·인터넷 분야의 분류체계를 비교·분석하여 새로운 분류체계 모형을 제시하였다. 1998년에는 최희윤(1998)이 인터넷 정보자원의 분류체계를 DDC와 비교 분석하여 분류체계 설계시 고려할 사항을 언급하였다. 또한 같은 해 남영준(1998)도 기존의 웹 문서 분류체계에 대한 비교·분석을 통해 웹 검색엔진용 분류체계를 새롭게 설계하여 제시하였다. 최재황(1998)은 DDC 분류체계를 활용하여 문헌정보학분야 인터넷 학술정보자원에 대한 분류코드를 설계하였다.

3 주제분류체계의 이론적 배경

3.1 분류의 의미 및 원리

어떤 한 개념 또는 사건을 유사한 속성을 근거로 군집화하는 것을 분류라 하는데, 사전적 해석은 “종류에 따라 나누는 것”으로 간단하다. 이를 구체화한 해석으로 Oxford English Dictionary(1989)에서는 “어떤 사물을 그 대

2) 연구 초기에는 9개였으나 최종마무리 과정에서 “Aza”는 서비스 종단으로 제외하였음.

상이나 유사성에 의해 류별(Class)로 나누거나 배열하는 행위”라고 하였다.” 여기에서 “개념”이라는 용어의 의미는 추상적인 사고에 사용되는 어느정도 구체적 또는 추상적 정신작용, 생각, 관념, 사상 등으로 이성활동의 내용이다(이홍용 1997).

일반적으로 분류에는 자연적 분류와 인위적 분류가 있다. 자연적 분류는 자연현상의 객관적 성질을 근거로 하여 이미 존재하고 있는 유사한 속성을 발견하여 조직화하는 것으로, 원소의 주기율 예에서 주기율은 인위적이지 않은 화학적 원소들의 속성에 의한 것이다. 인위적 분류는 구분의 필요성에 의해 인위적 조직을 통해 구분하는 것으로, 일반사회에서 임의적으로 설정하는 일반분류와 학문 연구과정에서 얻어진 개념분석의 결과로 체계화한 학문분류로 구분된다. 도서관·정보센터 등에서 활용하고 있는 문헌분류체계도 이러한 학문분류체계를 바탕으로 하여 지식의 효과적 이용을 목적으로 구체성과 유용성을 강조한 넓은 의미의 학문분류라 할 수 있다(최정태 1998).

본 연구의 목적이 웹 문서의 디렉토리 구성을 위한 지침을 마련하는 것이므로 전제한 분류 유형으로 볼 때 인위적 분류, 특히 문헌분류체계의 분류원칙이 적용될 수 있다고 볼 수 있다.

3.2 분류체계의 종류 및 특징

인위적 분류체계의 예는 우리 주위에서 어렵지 않게 찾을 수 있다. 일상생활에서 볼 수 있는 비디오 샵의 구분으로부터 정부, 공공기관 등에서 사용하는 공문서 분류체계, 신문·방송 등의 언론·방송사의 분류체계, 지적내용을 전

반적으로 다루고 있는 특허정보의 특허청 분류 체계 등 해아릴 수 없을 정도로 다양하다.

또한 분류란 속성이나 내용의 유사성에 따라 군집화한 것으로 볼 수 있으며, 보편적인 속성 또는 유사성은, 논리적·지리적·연대·기원(유전적)·역사적(연대, 기원, 지리적 개념의 복합)·진화적(단순에서 복잡으로의 유사성 정도)·역동적·자모순·수학적(수치, 부호에 따른 것)분류 등 9가지 속성들이라고 지적한 바(Richardson 1930)와 같이 위의 인위적 분류체계 예들도 Richardson의 9가지 분류유형의 하나 또는 복합적인 결합이라고 볼 수 있다.

그런데 본 연구의 대상인 웹 문서 내용과 속성의 다양성으로 볼 때 많은 인위적 분류체계들 중 문헌의 내용과 속성을 근거로 군집화 할 수 있도록 조직된 문헌분류체계의 유형을 분석하여 참조함이 타당할 것으로 사료된다.

문헌분류는 주제에 따라 구분하고 형식에 따라 배열함을 원칙으로 한다. 따라서 문헌분류체계는 1차적으로 주제 접근을 전제로 함을 알 수 있다.

이러한 문헌분류법은 분류방식에 따라 주제 내용을 10구분씩 점진적으로 세분하는 십진식과 숫자와 문자를 단독 또는 혼합하여 사용하는 비 십진식이 있다. 문헌분류법의 구조원리상으로는 열거식, 분석합성식, 준열거식으로 구분할 수 있다.

이상의 문헌분류법 유형들 중 시스템 관리자들이 인터넷상의 웹 문서를 분류하고, 이용자들이 필요한 문헌에 접근하는데 적용될 수 있는 주제 분류체계로는 구조원리상 열거식 분류법이며 분류방식으로는 전개성이 우수하고 논리적 접근이 가능한 비십진식 분류체계가 타당

할 것으로 판단된다.

문헌분류체계 중 십진식의 대표적 분류법은 DDC로서 학문분류에 근거하여 이용의 효율성을 강조한 분류체계를 갖고 있다. 즉, 전 지식 분야를 10개의 류로서 확장함으로써, 지식의 분화과정과 변화에 적극적으로 대응하는데 한계가 있을 수 있다. 그러나 주제의 상하 계층관계를 논리적으로 전개함으로써 이용자가 주제 범위를 확인하면서 접근하는데 용이한 체계를 유지하므로 웹 문서 분류체계의 논리성을 적용할 수 있을 것으로 판단된다.

비 십진식의 대표적 분류법은 LCC를 지적 할 수 있다. 이 분류법은 주제의 구분과 배열에 진화의 순서를 따르는 원칙이 강조된 Cutter, C.A.의 EC(Expensive Classification)분류체계(최정태, 양재한, 도태현 1998)의 주류배열을 참고하였다. 이 분류법의 특징으로는 EC의 분류체계를 참고하되 미의회도서관(LC) 장서를 조직하기 위한 목적으로 LC 장서구성과 편 이성을 고려하였기 때문에 특히 과학·기술분야에 우수한 분류법으로 평가받고 있다.

위의 대표적 분류체계 분석을 통해서 본 연구목적에 부합하는 적용 요소를 살펴본다면, 우선 주류의 선정에 있어서 이용목적에 합당한 주제를 선정하고, 이용자인 인간의 문제와 사고영역을 고려하며, 각 주제배열에 있어 조기성을 살리고, 논리적 근거에 의해서 배열함이 필요하다 하겠다. 또한 전개성과 최신성 유지를 위한 계속적인 보완이 필요하다 하겠다.

3.3 주제어와 용어 관계

웹 문서 디렉토리 검색엔진의 분류체계는 기

본적으로 주제어의 상호관계(상·하, 연관관계)를 나타내고, 주제 접근점을 통제하는 통제언어 색인시스템을 따른다고 볼 수 있다. 이것은 전조합 색인시스템의 디소러스 작성원칙과도 유사한 것으로서, 각각의 디렉토리 검색엔진에서 사용하고 있는 주제어의 상하관계와 각 계층에서 사용하고 있는 용어의 연관관계 등을 참조할 수 있을 것이다. 따라서 주제어의 상호관계에 대한 검토는 필연이다.

용어통제는 근원적으로 이용자의 접근점을 안내하기 위한 목적을 갖고 있다. 따라서 주제명표목표나 디소러스는 색인시스템에서 이용자가 다양한 용어로 탐색할 때의 탐색실패를 최소화하기 위한 도구로서 사용하는 것으로서 탐색자가 사용할 가능성이 있는 용어를 망라적으로 포함하는 방대한 것이다. 또한 용어간의 계층관계, 동등관계, 연관관계 등을 밝힘으로서 불필요한 접근을 감소시키는 역할도 하고 있다.

이러한 용어의 망라성과 주제분야 용어의 상호관계, 도구의 방대함으로 용어통제를 위한 도구로서 현재 지식 전 분야를 수용하고 있는 주제명표목표나 디소러스는 그리 많지 않으며³⁾ 주제분야별 도구들을 개발하거나 기존의 도구들을 이용하고 있는 것이다.⁴⁾

이렇게 용어통제 도구들은 기존의 문헌정보 이용을 위해 주제어를 탐색 접근점으로 이용하는 시스템(전조합 또는 후조합색인)에서는 많은 역할을 하였으나, 웹 문서를 위한 주제분류

3) LC의 Subject Headings, Sears List of Subject Headings, 이재철의 주제명표목표 등이 있다.

4) 의학분야의 Mesh, 과학기술분야의 Spines Thesaurus.

체계 하에서의 역할은 주제분류를 위한 계층관계와 연관관계 의한 링크의 역할에 집중될 것이며, 따라서 주제명표목표 및 디소러스의 계층적 주제어를 참고하는 정도의 활용이 가능하리라 판단된다.

또한 현재 웹 문서의 딕렉토리 검색엔진들은 인터넷상에 등록되어 있는 모든 주제의 문서들을 대상으로 검색하며, 그를 위한 주제어 또한 모든 주제를 망라하고 있으나 용어통제를 위한 도구개발에서는 소홀하여 효율적인 검색이 되지 못함을 볼 수 있다.

따라서 효율적 정보검색을 위해서는 용어통제를 위한 도구개발이 우선되어야 하며, 용어통제도구 또한 일반주제와 특정주제로 특성화하여 개발하고 각 검색엔진들에 적용하는 것이 필요할 것으로 판단된다.

4 딕렉토리 검색엔진 분석

현재 국내에서 이용되고 있는 검색엔진 중 서론에서 밝힌 바와 마찬가지로 일반주제 딕렉토리 검색엔진을 접속하여 메인 메뉴의 류 개념을 나타내는 최상위 주제의 딕렉토리와 그 하위 딕렉토리를 2001년 4월 15일부터 5월 5일까지 실사하여 분석하였다. 분석은 각 엔진의 일반적 특성, 딕렉토리 구성의 분류체계, 계층 간 논리성 등에 대해서 조사하고, 특히 심마니, 한미르(정보탐정), 네이버, 야후코리아, Zip은 딕렉토리의 변화 형태를 분석하기 위해 1998년 12월 현재와 2001년 5월 현재의 딕렉토리를 비교하였다.

4.1 딕렉토리 검색엔진의 일반적 특성 분석

인터넷을 통해 유통되는 정보는 양적으로도 통계가 정확하지 않을 정도로 많으며 또한 정보제공자에 의해 통제되지 않았을 뿐만 아니라 질적으로도 검증되지 않은 것이 많다(Floridi 1996). 이러한 이유로 정보이용자가 필요한 정보를 검색도구의 도움 없이 탐색하기란 거의 불가능에 가깝다 하겠다. 따라서 인터넷의 정보 찾기는 탐색도구인 검색엔진의 이용이 필연이며, 정보의 종류와 형태, 내용이 다양하여 검색엔진의 성능이나 탐색방법도 다양하다.

일반적으로 검색엔진은 키워드 탐색형, 주제별 딕렉토리 탐색형으로 대별되며, 그 외 딕렉토리와 키워드 시스템을 혼합한 혼합형이 있고, 자체 DB 외에 여러 탐색엔진을 참조하여 검색하는 메타검색엔진 등이 있다.

이러한 검색엔진을 색인시스템으로 보면 키워드탐색형은 전장에서 설명한 후조합색인에 해당되며, 주제별 딕렉토리형은 전조합색인에 해당한다 하겠다. 그러므로 본 연구의 목적인 딕렉토리 검색엔진의 딕렉토리 구성에서는 전조합색인 시스템의 특성인 색인항목의 접근성을 참조하고, 온라인상의 불 가시적인 탐색시스템의 특성을 고려하여야 할 것이다. 따라서 본 일반적 특성 분석에서는 류 개념의 분류체계와 류 개념과 하위 개념과의 포함관계를 나타내는 계층관계를 제외한 일반적 사항을 대상으로 분석하였는데 즉, 딕렉토리 항목 선정, 색인 항목의 일관성, 인터페이스, 문서수 제공 여부와 일치도, 하위계층 표목 확인 가능 여부, 주제어의 표기, 배열방법 등을 조사하였다.

또한 디렉토리 검색엔진은 학문분류와 논리적 체계에 기초한 문헌분류법과는 달리 분류체계와 하위주제의 개정이 수시로 이루어지는 특성을 지니고 있으므로, <표 1>에서 보는 바와 마찬가지로 새로운 분야의 삽입 및 통·폐합이 역동적으로 이루어지고 있음을 볼 수 있다.

색인항목에 있어서도 체계적인 주제명표목표 및 선택원칙에 의하기보다는 웹 문서상의 모든 용어를 색인 대상으로 하고 있으므로 일관성과 논리성이 결여되어 있음을 알 수 있다.

인터넷페이지 또한 다양하여 문서 수 제공여부, 하위계층의 완전제공여부, 확장·제한검색수단 제공여부 등이 다양하여 이에 대한 기본적인 표준화가 필요함을 알 수 있다.

즉, 각 주제에서 링크된 문서 수를 제공하는 것 중 최상위주제에서 문서수 또는 하위 주제 수를 알 수 있는 것은 전혀 없고, 하위주제에서는 까치네를 제외한 전체가 제공하였는데, 제시된 링크 수와 실제의 문서 수가 일치한 것은 심마니, 네이버이며 기타 대부분은 부분적으로 일치하였다. 또한 링크된 문서 수와 하위 디렉토리 수를 나타내는 방법이 일정하지 않아⁵⁾ 이 용자는 그 의미를 알기가 어렵다. 이것은 이용자가 새로운 주제 또는 상·하위주제로의 접근여부를 결정하는 중요한 요소가 될 수 있다.

하위계층 주제의 제공여부도, 계층구조가 명확하지 않고 역동적 특징을 보이는 디렉토리 검색엔진에서는 해당주제의 범주를 판단하는 근거가 되므로 중요한 요소라고 생각된다. 그러나 조사 대상 모두가 하위주제의 내용을 볼 수 없었다. 이것은 일반적으로 배열할 하위주제가 많아서 한 화면으로 제공할 수 없는 이유로 판단된다. 이러한 경우 주제의 상하관계를

볼 수 있는 기능이 필요하다 하겠으며, 하위주제의 주제어 수를 제시하는 방법도 고려해볼 수 있다고 판단된다.

주제어의 표기에서도 복합주제를 “/”, “,”, “와(과)”, 빈칸 등 다양하여, 하위개념 또는 한정적 개념의 의미와 복수주제의 개념이 구분하기 어렵게 되어 있음을 알 수 있다.

주제의 배열에 있어서는 심마니, 야후코리아, ZIP 이 자모순 배열방법을 채택하고 있었고, Search-korea는 분야별 Top10, 추천사이트, 개인 페이지를 제외한 모든 주제를 자모순으로 배열하였다. 이것은 검색엔진들의 주제선정이 논리적 근거에 의하지 않은 것으로 평가되는 점을 고려할 때, 당연히 자모순 배열이 되어야 이용자가 접근하기에 수월하다고 볼 수 있다.

1998년의 디렉토리와 2001년의 그것을 비교 할 때, 주제 명을 전혀 바꾸지 않은 것은 야후 코리아와 Zip이었고, 다른 검색엔진은 주제통합, 분리, 신설, 제거 등⁶⁾을 하여 변화를 꾀하였으나, 류 개념의 최 상위주제에는 선택에 신중을 기해야 하며, 한번 선택한 주제에는 가급적 그대로 유지함이 필요하리라 판단된다. 또한 서론에서도 밝힌 바와 마찬가지로 검색엔진의 변화에 있어서 일반주제 검색엔진들로부터 점차 전문주제 검색엔진으로의 이행이 예견되며, 같은 맥락에서 법률, 종교 등의 주제 명은

5) 하위디렉토리에서 “(??)”로 표시한 것은 “디렉토리 수/링크 된 사이트 수”로 표시한것과 “(00)”으로 하위 디렉토리, 또는 사이트 수를 표시하는 등 다양하다.

6) 통합(네이버 : 사회·문화, 정부·공공기관을 사회·문화로), 분리(한미로 : 학문을 과학기술, 인문사회과학으로), 신설(심마니 : 게임, 쇼핑 등), 제거(심마니 : 개인홈페이지) 등의 예가 있다.

전문주제 검색엔진에서 다루어지는 사례가 증가하였기 때문에 심마니에서는 법률, 종교가 주제 명에서 제거된 것으로 판단된다.

4.2 디렉토리 검색엔진의 분류체계 분석

이미 선행연구에서 지적한 바와 같이 온라인 환경에서도 기존의 분류체계를 이용하는 것이 검색효율성을 높여준다는 연구결과가 있듯이, 현행 디렉토리 검색엔진 중에는 DDC를 적용한 Mr. Dui's Topic Finder와 Yahoo, UDC를 적용한 GERHARD, LCC를 적용한 The WWW Virtual Library 등이 있다. 이들 검색 엔진들은 류 개념에서 각각 적용한 문헌분류체계를 전체 또는 일부를 적용하고, 하위계층에 서는 온라인 시스템상의 특성과 제한점을 고려하여 분류 항목 수를 조절하여 색인시스템의 비대함을 함축적으로 조정 제공하는 특성을 갖고 있다.

분석대상 검색엔진으로 국내는 심마니, 한미로(정보탐정), 야후코리아, 네이버, 까치네, Lycos-korea, Search-korea, Zip 등이다.

디렉토리 검색엔진의 분류체계는 학문분류를 기초한 문헌분류체계를 따르고 있는 것과 그렇지 않은 것으로 대별되며, 전 주제분야를 대상으로 동일한 비중으로 분류하는 것과 특정 분야를 중점대상으로 하는 것으로 구분된다.

본 연구의 대상으로 선정한 검색엔진들은 모두 전 주제분야를 대상으로 한 것으로서 최상위 주제를 12~15개로 선정하고, 하위 주제로 57개(네이버)까지 선정하고 있다.

선행연구에서 살펴본 바와 마찬가지로 디렉토리 구성에 있어서 문헌분류체계의 도입이 필

요하며 또한 유용성이 있음을 밝힌 바 있다. 그려한 배경에 따라 분석대상 디렉토리 검색엔진의 최상위 주제와 문헌분류체계인 DDC의 류 개념과의 일치성을 비교하면 다음과 같다.

즉, 디렉토리 검색엔진의 최상위 개념과 DDC의 류 개념이 일치하는 것은 종교, 사회과학, 자연과학, 응용과학, 예술 등이고, DDC의 강 개념과 일치하는 것은 공공기관, 뉴스·미디어·언론, 정치, 경제, 법률(학), 사회(문제), 교육, 의학, 레크레이션, 레저·스포츠, 오락·취미, 지역정보 등이다.

이 중 류 개념에서는 한 개가 일치하는 것은 심마니, 네이버, Lycos-korea 등이고, 2개가 일치하는 것은 Search-korea, 3개가 일치하는 것은 야후코리아, Zip 등이고, 4개는 까치네, 한미로 등이다. 또한 강 개념에서는 조사 대상 모두가 5개 이상 일치하였는데, 네이버, 야후코리아, 까치네가 6개, 심마니, 한미로, Search-korea가 각각 7개, Zip이 8개가 일치하여 1998년에 비해 월등히 많은 수를 최상위 개념으로 채택하였다. 류 개념보다 더 상위 개념으로는 학문, 과학(네이버, 심마니, Lycos-korea, Search-korea) 등이다.

또한 <표 2>에서 보는 바와 같이 분석대상 검색엔진 간에 최상위 계층에서 중복하여 출현하는 주제 중 경제·비지니스가 가장 많은 검색 엔진에서 제공되는 주제로서, 완전하게 동일한 명칭과 유사한 명칭 또는 부분일치 등을 합하여 총 9개 엔진에서 제공되며, 그 다음이 교육, 뉴스·미디어·언론, 사회·문화, 컴퓨터·인터넷 등이 8개, 레크레이션·스포츠, 문화·예술, 건강·의학 등이 6개 등으로 나타났다. 이것은 분석대상인 검색엔진(국내 중심)의 최상

위 주제가 각기 천차만별 식으로 다양하여 이용자가 찾고자 하는 주제를 예측하기가 어려움을 대변하고 있는 것이다. 이 결과는 1998년도의 디렉토리와 비교해 볼 때 순위의 변동이 심한 것을 볼 수 있는데, 이러한 결과를 보더라도 현행 디렉토리 검색엔진들의 분류체계는 극히 비논리적, 비체계적이고 매우 다양하다는 것을 알 수 있다.

남영준(1998)은 해당주제에 링크 된 문서 수를 근거로 분류한다고 하였으나, 최재황(1998)은 온라인 환경의 분류체계에 있어서 문헌분류체계인 DDC의 적용이 유용함을 여러 선행연구 사례를 통해 입증하고 있다. 다만 인터넷 정보의 유동적 특성에 따라서 하위계층의 주제는 링크 된 문서 수에 따라 역동적으로 조

절할 필요가 있다고 사료된다.

4.3 디렉토리 검색엔진의 계층관계 분석

검색엔진들의 계층관계에 있어서, 우선 최상위 계층(류 개념)과 하위계층(강개념)의 주제어 수를 비교하고, 다음으로 최상위계층의 주제어 유형을 비교한다. 그리고 동일한 명칭의 최상위 주제의 하위계층을 표본으로 추출하여 비교하고자 한다.

첫째, 계층간 주제어 수의 비교로서, <표 1-1>과 <표 1-2>에서 보는 바와 마찬가지로 국내 검색엔진의 최상위 주제어는 12개로부터 15개로 평균 약 14개이며, 8개의 검색엔진 전체에서 출현하는 주제 수는 62개로 평균 약 6개의

<표1-1> 디렉토리검색엔진의 최상위 주제 명 및 하위 주제 수('98, '01비교)

검색엔진	류개념의 주제명(1998)		류개념의 주제명(2001)		증감
한미르	건강, 병원	8	건강, 의학	37	29
	학문	2	과학기술	32	
			인문사회과학	27	
	교육, 출판	7	교육, 취업	22	15
	생활, 주택	3	생활, 가정	23	20
			정부, 공공기관	13	8
	정치, 행정, 법	5			
	컴퓨터 인터넷	13	컴퓨터, 인터넷	24	11
	홈페이지	2	취미, 게임 홈페이지	45	43
	영화, 음악, 연예	4	엔터테인먼트	28	24
	예술, 문화, 출판	8	문화, 예술	25	17
	언론, 매체	3	뉴스, 미디어	24	21
	경제, 산업	7	비지니스, 경제	25	18
	취미, 레저	4	여행, 레저, 스포츠	9	5
			지역정보	11	
			사회, 종교	18	
계		66		363	297

〈표1-1〉 디렉토리검색엔진의 최상위 주제 명 및 하위 주제 수(98, '01비교)

검색엔진	류개념의 주제명(1998)	류개념의 주제명(2001)	증 감	
심마니	건강, 병원	23	건강, 의학	30 7
	교육, 대학	71	교육, 학습	29 12
	과학	8	학문, 참고자료	18
	역사	6		
	산업, 경제	11	비지니스, 경제	32 21
	법률	6		
	정치, 행정	7	정치, 행정	22 15
	뉴스, 언론매체	5	뉴스, 언론	17 12
	사회, 문화	24	사회, 생활	27 3
	어린이, 청소년	9		
	종교	6		
	컴퓨터, 인터넷	12	컴퓨터, 인터넷	23 11
			엔터테인먼트	23
	오락, 취미	44	취미, 스포츠	36 -20
	스포츠	12		
	예술	13	예술	23 10
	개인홈페이지			
계			게임	12
			쇼핑	45
			지역정보	14
		203		351 148
야후 코리아	건강, 의학	40	건강, 의학	46 6
	교육	26	교육	30 4
	뉴스미디어	33	뉴스미디어	31 -2
	레크리에이션, 스포츠	20	레크리에이션, 스포츠	23 3
	비지니스, 경제	32	비지니스, 경제	30 -2
	사회과학	34	사회과학	33 -1
	사회, 문화	28	사회, 문화	30 2
	엔터테인먼트	29	엔터테인먼트	31 2
	예술, 인문	24	예술, 인문	28 4
	자연과학	47	자연과학	20 -27
	정부	15	정부	19 4
	지역	4	지역	4 0
	참고자료	27	참고자료	30 3
	컴퓨터, 인터넷	37	컴퓨터, 인터넷	44 7
계		396		399 3

〈표1-1〉 디렉토리검색엔진의 최상위 주제 명 및 하위 주제 수('98, '01비교)

검색엔진	류개념의 주제명(1998)		류개념의 주제명(2001)		증감
네이버	건강, 의학	26	건강, 의학	39	13
	뉴스, 미디어	21	뉴스, 미디어	30	9
	레크리에이션	14	레크리에이션	26	12
	사회, 문화	29	사회, 문화	39	10
	정부, 공공기관	12			
	사업, 경제	21	사업, 경제	28	7
	지역정보	5	지역정보	30	25
	컴퓨터	26	컴퓨터, 인터넷	37	11
	사회과학	17			
	인문, 예술	17	학문, 과학	30	-31
	자연과학	27			
	교육	22			
	참고자료	14		4	-32
	연예, 오락	20		25	5
ZIP			가정, 여성	24	
			게임	25	
			쇼핑	38	
			스포츠	57	
	계	271		432	161
	건강, 의학	26	건강, 의학	25	-1
	교육	24	교육	26	2
	뉴스, 미디어	21	뉴스, 미디어	30	9
	레크리에이션	14	레크리에이션	17	3
	사업, 경제	21	사업, 경제	26	5
계	사회과학	17	사회과학	27	10
	사회, 문화	29	사회, 문화	30	1
	연예, 오락	20	연예, 오락	19	-1
	인문, 예술	17	인문, 예술	19	2
	자연과학	27	자연과학	30	3
	정부, 공공기관	12	정부, 공공기관	13	1
	지역정보	5	지역정보	5	0
	참고자료	14	참고자료	16	2
	컴퓨터	25	컴퓨터	25	0
	계	272		308	36

〈표 1-2〉 디렉토리검색엔진의 최상위 주제명 및 하위 주제수(2001년)

Search-Korea		LYCOS-KOREA		까치네	
류 개념의 주제명	하위 주제수	류 개념의 주제명	하위 주제수	류 개념의 주제명	하위 주제수
개인홈페이지	19	건강, 의학	24	건강	14
건강, 의학	25	과학, 학문	28	교육	17
경제, 기업	17	교육, 참고자료	33	뉴스, 미디어, 언론	14
교육, 학교	24	뉴스, 미디어	9	레크레이션, 스포츠	16
문화, 예술	12	레크레이션, 스포츠	23	법, 법률	7
사전, 자료	11	문화, 사회	26	사회, 문화, 생활	18
사회, 정치, 종교	23	비지니스, 경제	35	산업, 경제	20
쇼핑, 생활	13	생활, 가정	17	연예	17
신문, 방송	9	쇼핑	27	예술	20
여행, 스포츠	28	예술, 엔터테인먼트	18	인문, 사회과학	16
연예, 오락	12	취업정보	20	자연과학, 공학, 의학	26
지역, 국제정보	9	컴퓨터, 인터넷	16	정부, 공공기관	16
컴퓨터, 인터넷	20			종교	11
학문, 과학	53			컴퓨터, 인터넷	25
계	275		276		237

주제들이 중복해서 출현한다고 볼 수 있다. 또 첫 번째 하위 주제어는 4개로부터 57개로 평균 약 24(1998년은 15개)이며 편차가 매우 큰 것으로 보아 각 검색엔진들의 주제 구성이 매우 다른 것으로 평가할 수 있겠다. 둘째, 최상위 계층의 주제어 분석으로, 앞에서 보는 바와 같이 국내 검색엔진은 평균 14개의 최상위 개념어를 제공하고 있는데, 표기와 범위에 있어 약간의 차이를 보이고 있으나 많은 주제어가 일치하고 있었다. 특히 네이버, 야후코리아, Zip은 거의 일치하고 있음을 알 수 있다(〈표2〉참조).

또한 주제어의 일치도를 보면 2개 이상의 검색엔진에서 일치하는 주제는 18개로서 분석대상 국내 검색엔진(62개)의 약 29%(1998년도는 약 20%)인데, 유사주제(33개)를 포함하면 약82%에 이른다. 이러한 분석결과를 기초로

하여 디렉토리 검색엔진에서 최상위 주제로 선택하는 기준을 정한다면 이용자의 주제 접근상의 혼란을 방지할 수 있으리라 판단된다.

셋째, 계층간 논리성 분석으로, 최상위 주제 하의 하위계층 주제어가 논리적으로 포함관계에 있는지를 분석하고자 하였다. 본 논리성 분석에서는 역시 기존의 분류체계인 KDC를 준거하였는데, 이것은 앞에서도 밝힌 바와 같이 기존의 문헌분류체계는 학문분류를 기초로 하였으므로 주제의 범주에 대한 계층의 논리성이 우수하기 때문이다. 논리성 분석에서는 분석 대상 검색엔진의 전부, 전 주제를 분석하는 것은 너무 방대하여 국내검색엔진의 다빈도 출현 주제인 “교육”부분을 대상으로 하였다.

〈표 3〉에서 보는 바와 같이 조사 대상 검색엔진들은 최소 12%(17개 주제중 2개)부터 최대

33%(24개 주제 중 8개)까지 KDC의 강 부분과 일치하여 전체적으로 약 22.3%정도의 일치성을 보이고 있다. 또한 상위 개념인 교육학을 주제어로 선정한 경우는 까치네, 한미르를 제

외한 기타 검색엔진이 해당되며, 기업간 거래(야후코리아), 뉴스와 미디어(야후코리아), 언론·방송(Search-korea), 과학(Zip) 등 명확하게 다른 분야이거나 최상위 주제명

〈표 2〉 최상위 계층 주제 분포

동일한 명칭			유사 또는 부분 일치 명칭	
주제어	검색엔진	엔진수	주제어(검색엔진)	엔진수
건강, 의학	심, 한, 네, 야, Z	5	건강(까)	1
게임	심, 네	2		
교육	야, Z, 까	3	교육 · 참고자료(네, L), 교육 · 취업(한), 교육 · 학교(S), 교육 · 학습(심)	5
뉴스, 미디어	한, 네, 야, Z, L	5	뉴스 · 미디어 · 언론(까), 뉴스 · 언론(심), 신문 · 방송(S)	3
레크레이션 · 스포츠	야, 까, L	3	레크레이션(네, Z), 스포츠(네)	3
문화 · 예술	한, S	2	예술(심, 까), 예술 · 인문(야, Z)	4
비지니스 · 경제	심, 한, 네, 야, L	5	사업 · 경제(Z), 산업 · 경제(까), 정치 · 경제(심), 경제 · 기업(S)	4
사회 · 문화	네, 야, Z, L	4	사회 · 문화 · 생활(까), 사회 · 생활(심), 생활 · 가정(한, L)	4
사회과학	야, Z	2		
쇼핑	심, 네, L	3	쇼핑 · 생활(S)	1
연예, 오락	Z, S	2	연예(까)	1
예술, 인문	야, Z	2	예술(심, 까)	2
엔터테인먼트	심, 한, 야	3	엔터테인먼트 · 예술(네, L)	2
자연과학	야, Z	2	자연과학 · 공학 · 의학(까), 과학 · 기술(한)	2
정부 · 공공기관	한, 까, Z	3	정부(야)	1
종교	까	1	사회 · 정치 · 종교(S), 사회 · 종교(한)	2
지역정보	심, 한, 네, 야, Z	5	지역 · 국제정보(S)	1
컴퓨터 · 인터넷	심, 한, 네, 야, 까, Z, L	7	컴퓨터(Z)	1
학문 · 과학	네, S, L	3	학문 · 참고자료(심), 참고자료(야, Z), 사전 · 자료(S)	4
가정 · 여성	네	1		
법 · 법률	까	1		
여행 · 레저 · 스포츠	한	1	여행 · 스포츠	1
취미 · 개인홈페이지	한	1	개인홈페이지(S), 취미 · 스포츠(심)	2
취업정보	L	1		1

* 약어별례 : 심마니(심), 한미르(한), 네이버(네), 야후코리아(야), 까치네(까), Search-Korea(S), Lycos-Korea(L), ZIP(Z)

〈표 3〉 KDC와 검색엔진간의 교육부분 하위계층 비교

검색엔진	주제명	강 개념 하위주제수
까치네	유아·조기교육, 특수교육	2/17 (12%)
네이버	교수기술(교육방법), 대학(교), 중·고등학교, 초등학교, 성인교육·평생교육(사회교육), 유아교육, 특수교육	7/41(17%)
심마니	교육행정, 대학교육, 유아교육, 중·고등교육, 초등교육, 특수교육, 평생교육(사회교육)	7/29 (24%)
야후코리아	교수법(교육방법), 유아교육, 주제·과목별(교육과정), 특수교육, 평생교육(사회교육), 행정·공공기관	6/30 (20%)
한미르	대학교, 성인·평생교육(사회교육), 유아·조기교육, 초중고교육, 특수교육	5/22 (23%)
Lycos-Korea	대학·대학원, 유아·조기교육, 중등교육, 초등교육, 특수교육, 평생·성인교육사회교육	6/33 (18%)
Search-Korea	고등학교, 대학(교), 대학원, 유아교육, 중학교, 초등학교, 특수교육, 평생교육·재교육(사회교육)	8/24 (33%)
Zip	고등학교, 대학, 중학교, 초등학교, 유아교육, 교수기능(교육방법), 특수교육, 평생교육(사회교육)	8/26 (31%)

* 네이버, 심마니, 야후코리아, Lycos-Korea, Search-Korea, Zip 등은 류 개념의 “교육학”을 하위주제인 강 개념에 분류하고 있음.

들이 포함되어 있었다.

이와 같은 주제어 선정 현황을 볼 때 검색엔진들의 주제 선택은 검색엔진 개발자들의 자의적인 판단에 따라 정한 것으로 볼 수 있는 것으로 디렉토리의 계층성에 있어서 비논리적인 부분이 많다고 볼 수 있다.

4.4 분석결과에 대한 논의

첫째, 일반적인 특성에서 주제선정, 배열, 인터페이스, 주제어 표기, 배열방법, 각 주제어에 링크된 문서 수 등의 요소를 비교하였는데, 각 검색엔진들은 이용자들의 위한 배려가 부족하고, 수집대상 문서가 동일함에도 불구하고 많은 요소들이 각기 다른 형태를 보이고 있었다.

특히 하위 디렉토리에서의 링크 된 문서 수 또는 그 하위 디렉토리 수를 표기하는 방법이 동일하지 않고, 동일한 검색엔진에서조차 일정하지 않아서⁷⁾ 키워드 방식이나 메타검색엔진의 자동색인 방식의 부정확성을 연상하게 하고 있었다. 이는 디렉토리 검색엔진은 다른 검색엔진과는 달리 웹 문서(또는 사이트)의 수집과 해당 주제에 대한 배정과 더불어 그 결과를 운영자들이 항상 점검해서 수정 보완해야만 신뢰도를 높일 수 있다고 판단된다.

7) 한 예로 네이버의 최 상위주제 “스포츠” 하위의 “스포츠개인@(62)”은 그 하위 디렉토리가 7개, 링크된 사이트가 5개로, “스포츠신문@(9)”는 신문 사이트만 9개, “축복(2)”은 사이트 링크만 2개 등으로 되어 있어서 “@” 표시 및 () 가 의미하는 것을 알 수 없었다.

둘째, 분류체계 분석에서는 주제선정의 논리성 면에서 기준의 문헌분류체계를 주로 따르지 않고 있음을 알 수 있는데, 이는 인터넷 정보의 다양성과 복잡성을 고려한다 해도 이용자가 인식하는 주제의 범주가 기준의 문헌분류체계의 주제를 연상하는 경향(신동민 1988)을 보이고 있으므로 주제선정에 동 분류체계들을 참고함이 필요하다고 판단된다. 다만 분석대상 검색엔진들의 일부는 DDC의 류 개념과 강 개념에서 일치하는 주제를 선정하였음을 확인할 수 있었다.

셋째, 계층관계 분석에서는 상·하위 계층간의 주제어 수, 각 검색엔진별 최상위계층 주제어의 일치도, 계층간 주제어의 논리성 등을 분석하였다. 즉, 계층간 주제어 수는 최상위 주제어 수에 있어서 큰 차이는 없었으나(12~15개), 하위 주제 수는 큰 편차를 보였다(5~57개). 최상위 계층의 주제어 일치도에서 “건강”, “교육”, “뉴스·언론”, “스포츠·레크레이션”, “사회·문화”, “자연과학”, “컴퓨터·인터넷” 등의 주제가 여러 검색엔진에서 출현함을 알 수 있었다.

계층간 논리성에서는 KDC를 준거로 하여 여러 검색엔진에서 다빈도로 출현한 “교육”을 분석하였는데 전체적으로 하위주제의 일치도가 약 15% 정도로서 빈약함을 보였다.

의 표기, 주제어간의 관계 및 이용자 인터페이스 등에 있어 혼격하게 다양한 형태를 취하고 있어서, 이용자가 동일한 내용을 검색하기 위해 여러 검색엔진을 탐색할 때 어려움을 겪고 있다.

특히 분석대상으로 선정했던 검색엔진들뿐만 아니라 기타 대부분의 디렉토리 검색엔진들은 공통적으로 인터넷상의 모든 분야 정보를 대상으로 수집하여 분류하므로, 특정분야만을 선별적으로 검색하기를 원하는 이용자는 만족 할만한 검색결과를 기대할 수 없는 것이 현실이다.

또한 정보접근의 어려움에도 불구하고 이용자가 검색을 실시한 후의 결과에 대해서 ‘검색된 웹 문서들의 내용이나 접근상의 제한사항들을 사전에 알지 못하므로 많은 시간을 낭비하게 되는 문제도 현행 디렉토리 검색엔진들이 해결해야 할 문제라고 판단된다.

따라서 본 장에서는 제4장의 분석결과를 토대로 모든 디렉토리 검색엔진들에서 공통으로 채용하여 적용할 지침을 제시함으로써, 이용자가 원하는 정보를 찾기 위한 최소한의 노력으로 최대의 만족을 갖도록 하고자 한다.

5.2 디렉토리 검색엔진에 적용할 지침

① 분류체계의 조건에 관한 지침

분류체계의 구성에는 지식체계를 기반으로 하는 문헌분류체계의 구성원칙에 준하되, 검색엔진의 일반적 특성 및 분류체계의 분석결과와 분류대상이 온라인 상의 정형화되지 않은 정보라는 점을 고려하여 다음과 같이 기본적인 조건을 제시할 수 있다.

5 검색엔진에 적용할 디렉토리 구성지침

5.1 지침 개요

현행 디렉토리 검색엔진들을 제4장에서 분석한 바와 마찬가지로 주제를 나타내는 주제어

첫째, 유연성으로서 융통성이라고도 할 수 있는 것으로 최 상위주제를 제외한 하위 디렉토리에서 새로운 주제의 삽입과 유사주제의 통합 및 분리, 세분 등 웹 문서의 생성 및 소멸과 해당주제의 문서 밀집도에 연계하여 주제어를 조정할 수 있도록 해야 한다.

둘째, 용어의 조기성으로서 하위계층의 모든 부문, 또는 일부 부문에서 동일하게 나타나는 주제는 동일한 용어와 동일한 순서로 배열되어야 한다. 즉 홈페이지, 동일계층분야의 일반주제, 복합주제, 역사 등의 내용은 동일계층의 전반부에 배치하고, 이러한 조기성의 대상이 되는 분류대상을 선정하고 안내하도록 한다.

셋째, 주제어 범위의 명확성으로서 주제어로 선정한 용어가 나타내는 개념의 범위가 명확해야 한다. 특히 현행 디렉토리 검색엔진들의 주제범위가 서로 일치하지 않는 것은 주제어의 용어개념이 불명확함에 따른 결과다. 따라서 검색엔진들은 주제의 범위를 정함에 있어 기존의 분류체계를 참조해서 그 범위를 정하고 “scope note”를 마련하여 적용의 일관성과 명확성을 확보해야 할 것이다.

넷째, 배열의 일관성으로 검색엔진의 접근점은 문헌분류기준과 같이 공통성을 강조한 숫자기호체계가 아니라 문자기호체계인 주제어 자체가 배열기준이 되므로, 각 계층에서 주제어의 자모순 배열을 원칙으로 한다. 다만 조기성의 대상이 되는 요소는 일반주제어에 우선해서 자모순 배열을 따르도록 한다. 이 예로서 Search-korea는 모든 주제를 자모순으로 배열하고 있으나 “분야별 Top 10, 추천 사이트, 개인홈페이지”는 예외로 하고 있음을 볼 수 있다.

② 주제의 계층관계에 관한 지침

검색엔진의 계층성은 동일계층의 주제어 수에 영향을 미치는 요소로써, 이것은 온라인 상에서 주제어를 확인하는 이용상 특성으로 신중한 선택이 필요하다.

첫째, 검색엔진의 특성에 따라 주제의 계층관계를 결정한다. 즉 최재황(1998), 김영보(1997) 등의 연구에서 보는 바와 같이 모든 검색엔진이 모든 주제를 대상으로 하지 않으므로, 특정주제를 중점적으로 제공하는 검색엔진은 해당주제의 계층관계를 특정적으로 유지하도록 한다. 이러한 특정주제의 계층관계 유지는 이미 많은 검색엔진들이 특정분야(예: 종교, 법률 등)에서 운영되고 있는데, 특정 주제에 대한 분류체계의 이용이 필수적이라 사료된다.

둘째, 상하위 계층간의 계층성 유지에 있어서 적정 주제 수를 유지하도록 한다. 즉 상위계층과 포함관계에 있는 주제는 하위계층의 주제 수를 고려하여 세분하거나 통합하여 적절 수를 유지해야 한다.⁸⁾ 이 경우 최상위 디렉토리를 제외한 하위 디렉토리에서의 분류체계의 유연성과 융통성을 고려하여야 한다.

셋째, 계층별 단계에 대한 해당 주제의 계층관계를 볼 수 있는 인터페이스를 제공한다. 이것은 애후에서 최종 하위 디렉토리를 화면 상단에 제시하는 것과 같다.

③ 주제어 선정, 배열에 관한 지침

주제어의 형태는 검색엔진들에서 매우 큰 다양성을 보이고 있으므로 주제접근에 혼란을 초래하는 요소이다.

첫째, 복수개념의 주제어를 단일 표목으로 선정할 때 가급적 복수개념을 포함하는 상위주

8) 계층관계 분석에 의한 결과로 각 계층의 최대 주제수는 40개까지로 하는 것이 바람직하다.

제를 선택하여 선정하고, 상위주제 선택이 불 가능한 경우 복수주제의 표기를 “.”로 통일함을 원칙으로 한다.

둘째, 모든 주제어는 동일계층에서 주제어의 자모순 배열을 원칙으로 하되, 조기성에서 제기한 요소들의 주제어를 우선하고, 다음으로 일반 주제어를 배열한다.

셋째, 최상위 주제어의 선정은 현재 이용되고 있는 검색엔진들에서 공통적으로 사용하고 있는 주제를 우선 선정하고, 그 외 검색엔진의 특성에 따라 약간의 주제어를 별도로 선정하여 제공한다.

넷째, 동일한 주제하의 하위주제어 선정과 배열은 동일한 방법으로 통일한다. 그 예로서 “Yahoo”의 “Education > Higher Education > Colleges and Universities …”를 들 수 있다.

④ 주제어간 상호관계에 관한 지침

인터넷상의 웹 문서가 하나의 주제만으로 색인될 수 없음은 문헌분류의 이치와 같다. 따라서 검색엔진에서 제공되는 주제어는 검색결과에 따라 상·하위개념 또는 관련개념으로 검색의 확장, 제한을 할 수 있도록 주제어간 상호관계를 정의하고, 필요시 검색화면에서 상호관계를 볼 수 있도록 인터페이스를 제공하도록 한다. 또한 동일 주제명이 다른 계층에서 출현하는 경우도 볼 수 있도록 한다.

⑤ 주제의 일반성, 특정성에 관한 지침

검색엔진의 특성화를 위해서 검색엔진의 주제범위를 제시하도록 한다.

첫째, 주제의 일반성에 있어서 상대적으로 포괄적인 개념의 주제는 그 주제어의 개념적 범주를 제시하도록 한다.

둘째, 특정성에 대한 개념표현은 특정 개념

어를 선정하되, 적절한 용어선정이 어려울 경우 한정적 개념으로 사용한 한정적 용어는 후치하고 용어간에는 “/”로서 구분한다.

셋째, 특정성을 고려하여 특정분야를 특화한 검색엔진에서는 해당분야의 분류표를 고려하도록 한다(예 : 특허부문 특화검색엔진은 “국제특허분류(IPC)”를 참조).

⑥ 이용목적에 따른 접근방안에 관한 지침

전체 웹 문서에 고르게 분포되어 있는 주제, 또는 분야에 관한 정보이용에 있어서 전체 또는 해당 하위 디렉토리 범위 안에서만 검색을 원할 경우는 별도의 인터페이스(키워드검색을 위한 박스 제공 등)를 통해서 검색할 수 있도록 한다. 즉 홈페이지, 역사, 쇼핑, 광고·홍보, 행사, 회의 등 여러 분야에서 검색될 수 있는 정보는 해당주제에서 검색할 경우 해당주제 이하 범위 내에서만 검색할지, 아니면 전 주제분야로 확대할지를 지정하는 별도의 인터페이스(키워드검색을 위한 박스 제공)에 의해서 목적에 따라 검색되도록 한다.

⑦ 계층간 웹 문서의 수용범위(문서 수)에 관한 지침

이용자의 인터넷 이용 습관에서와 마찬가지로 일정 수 이상의 문서가 검색될 경우 검색된 문서를 열람하지 않는 경향이 있으므로 최하위 계층에서 수용할 문서 수의 한계는 다음에 따르도록 한다.

첫째, 최하위 계층의 문서수가 30건 이상일 때 주제를 계층내의 수평분리, 또한 하위계층으로 재분류하도록 한다. 이것은 이용자가 일반적으로 검색건수가 30건 이하일 때 확인하는 경향(신동민 1988)에 따른 것이다.

둘째, 최하위계층의 문서수가 너무 적을 때

수평통합, 또는 상위계층에 재분류한다.

⑧비 주제 부문에 대한 적용방안에 관한 지침
본 연구의 대상인 디렉토리 검색엔진은 근원적으로 주제를 합리적으로 계층적 관계를 유지하도록 하고 있다. 그러나 주제에 의한 정보의 탐색뿐만 아니라 비 주제부문 즉, 정보의 형태(음성, 동·화상, 멀티미디어 등) 정보의 접근 가능성(접근자격제한 등), 인터페이스의 유형 등도 정보탐색의 효율성에 영향을 주는 요인으로 이들에 대한 지침이 필요하다.

모든 검색엔진에 기본적으로 갖추어야 할 인터페이스의 내용을 정리하면 다음과 같다.

첫째, 주제어의 열람기능 제시; 주제어 입력박스를 제공하고 주제어를 중심으로 한 계층 관련 관계 및 자모순 Browsing의 기능을 제공한다.

둘째, 모든 주제어에 문서건수 제시; 상·하위계층을 막론하고 모든 계층, 모든 주제에 색인된 문서건수를 제시한다.

셋째, 각 계층 및 주제분야에 분산된 특정주제(⑥항 참조)를 한곳에서 볼 수 있는 기능을 제공한다.

넷째, 검색 시 원하는 정보의 형태를 제한하는 기능을 제시할 수 있도록 한다.

⑨ 주제어 검색 및 feedback에 관한 지침

첫째, 이용자가 keyword box에서 접근을 시도하는 단어가 주제어 목록에 없을 경우 시스템에서 기록을 유지하고 일정횟수 이상 반복적으로 접근이 시도된 단어를 주제어로 선정하여 분류하도록 한다. 이것은 검색엔진의 주제어 체계를 모르는 이용자가 주제어의 계층 또는 자모순 열람을 하지 않거나 모르는 경우에 해당할 것이며, 빈번하게 접근하는 단어는 주제

어로서의 역할을 할 수 있다고 보기 때문이다. 또한 이 경우 디스크립터로 연결하기 위한 비디스크립터의 디스크립터로의 연결기능도 있도록 한다.

둘째, 각 계층에서 제공되는 메뉴의 주제어를 클릭하지 않고 keyword 입력박스에 직접 주제어를 입력할 경우 주제어를 열람기능으로 자동 전환되어 계층별, 자모순 등의 여부를 인터페이스에서 설정하도록 한다.

⑩ 유지관리에 관한 지침

현재 인터넷상에서 검색되는 문서의 구성 내용은 형식과 질에 있어서 너무나 다양하고 차이가 크다. 또한 문서내의 용어의 적절성과 접근제한 등의 다양성으로 자동분류의 한계가 있고, 급속한 변화로 인한 변경이 수시로 이루어 지므로 이에 대한 확인이 필요한 것이다. 따라서 다음과 같은 노력이 필요하다 하겠다.

첫째, 문서의 최종적인 분류와 주제할당은 자동분류와 더불어 관리자가 해당문서를 직접 접근, 확인하는 방법을 병행한다. 이는 현재 문헌분류에 있어서 문헌의 형식, 내용구성, 체제 등이 비교적 일정함에도 불구하고 모든 분야의 분류가 자동적으로 이루어지고 있지 않음을 볼 때 자동분류에 의한 주제할당은 한계가 있으며, 이로 인한 검색효율의 저하가 초래된다고 할 수 있기 때문이다.

둘째, 문서의 자동 Check에 의한 변동사항을 수시로 점검하여 이를 반영한다.

셋째, 웹 문서 제공자 및 가공자와의 협력체계를 구축하고, 내용수정 등의 정보를 수시로 보완하도록 한다.

넷째, 일정건수 이상 특정주제에 분류 또는 할당될 때 분류체계를 정기적으로 수정, 보완한다.

6 결론 및 제언

인터넷 정보검색 수단으로서의 검색엔진은 인터넷 공간에서 정보수집을 위해서는 필수요소라 할 수 있다. 현재 여러 유형의 검색엔진들이 포털기능을 갖고 운영되고 있는데, 인터넷 정보의 검색서비스 제공이라는 본래의 목적에 충실해서 이용자가 이용하기에 편리할 뿐만 아니라 검색결과의 효율성도 높이도록 해야 할 것이다. 그런데, 이용자들은 검색엔진들의 다양성과 성능의 차이에 의한 검색 결과의 비효율성으로 인해서 인터넷 공간에서의 정보 수집에 많은 어려움을 겪고 있는 것이 사실이다. 따라서 인터넷 검색엔진들은 무한경쟁시대에 있어서의 생존을 위해서도 계속적인 수정과 보완을 통해 이용자로 하여금 많이 이용할 수 있도록 해야 할 것이다.

이에 본 연구는 현행 디렉토리 검색엔진의 검색 효율성과 이용 편리성을 높일 수 있도록 하는 디렉토리 구성에 대한 연구를 통해 디렉토리 구성 지침을 마련하고자 하였다. 이를 위해 디렉토리 검색엔진들의 주제접근에 대한 편의성, 주제선정의 논리성 및 계층성, 명확성 등을 중심으로 분석하였다.

분석결과로서 현행 디렉토리 검색엔진들은 주제선정에서 논리성이 미약하고, 이용자에 대한 주제접근 편의성에 대한 배려가 부족하여,

각 검색엔진들이 다양한 형태의 인터페이스 및 문서분류체계를 유지함으로써 디렉토리 검색엔진으로서의 주제분류 체계에 대한 개선 방안이 필요함을 확인하였다. 따라서 본 연구에서는 디렉토리 검색 엔진의 이용자 접근점이 되는 디렉토리 구성에 적용할 지침으로 분류체계 구성 조건을 포함한 10가지를 제시하였다.

본 연구는 디렉토리 검색엔진의 디렉토리 구성에 관한 지침을 제시하였는데, 이 연구결과는 향후 모든 디렉토리 검색엔진에서 적용할 수 있으리라 판단된다. 다만 인터넷상의 정보유통은 그 양이나, 형태, 종류 등이 일반인의 상상을 초월할 정도로 빠르고 다양하게 변화하고 있으므로, 계속적인 보완을 위해서 후속 연구 및 개정이 필요하다고 판단된다.

또한 현재 이용되고 있는 많은 디렉토리 검색엔진들은 대부분이 전 주제분야를 대상으로 하고 있는데, 향후 특정주제분야를 특성화하여 주제를 세분화하고 전문화하여 해당 주제분야의 검색효율을 높일 수 있도록 하는 전문주제분야의 검색엔진이 많이 출현하여야 하며, 그런 검색엔진의 개발시 주제의 전문성과 특성성을 고려하여 주제전문가와 주제전문 분류체계의 참조가 필수적이라 하겠다. 이것을 위해서도 특정주제분야 웹 문서 분류체계 작성에 대한 계획적 연구가 필요하리라 판단된다.

참 고 문 헌

- 김영보. 1997.『인터넷탐색엔진의 분류체계에 관한 연구 : 컴퓨터·인터넷 분야를 중심으로』. 석사학위논문, 성균관대학교.
- 남영준. 1998. 웹 문서 분류체계의 설계.『제5회 한국정보관리학회 학술대회논문집』, 1997년 8월 19-20일. [서울: 한국정보관리학회]. 183-188
- 남영준. 1998. 웹 문서 분류체계의 분석 및 새로운 설계.『한국문현정보학회지』, 32(3): 207-230
- 신동민. 1988.『Web 데이터베이스의 효율적인 검색을 위한 분류체계 연구: 최종 보고서』. 한국데이터베이스진흥센터.
- 여찬기 등저. 2000.『공공기관의 홈페이지 구축·운영지침서』. 서울: 한국전산원.
- 이홍용 편저. 1997.『(객관식)분류론 : KDC4 판 수록 요점정리 및 해설』. 서울: 도서출판훈민.
- 정영미. 1997.『지식구조론』. 서울: 한국도서관협회.
- 최재황. 1998. 인터넷 학술정보자원의 디렉토리 서비스 설계에 있어서 DDC 분류체계의 활용에 관한 연구.『정보관리학회지』, 15(2).
- 최정태, 양재한, 도태현 공저. 1998.『문헌분류의 이론과 실제』. 부산: 부산대출판부.
- 최희윤. 1998. 인터넷 정보자원의 조직을 위한 분류체계에 관한 비교분석.『제5회 한국정보관리학회 학술대회논문집』, 1998년 8월 19-20일. 서울: 한국정보관리학회. 177-182
- 홍종남. 1997.『인터넷 정보검색엔진 256배 활용』. 서울: 영진출판사.
- Beghtol, C. 1998. "Knowledge Domains: Multidisciplinarity and Bibliographic Classification Systems." *Knowledge Organization*, 25(1/2): 1-12
- Dahlberg, Ingtraut. 1995. "The Future of Classification in Libraries and Networks, a Thoretical Point of view." *Cataloging & Classification Quarterly*, 21(2): 23-35
- Floridi, L. 1996. "The Internet : Which Future for Organized Knowledge, Frankenstein or Pygmalion? Part 1&2." *The Electronic Library*, 14(1): 43-48
- Hunter, Eric J. 1998. *Classification Made Simple*. Aldershot, Gower.
- Kumar, Krishan. 1981. *Theory of Classification*. 2nd rev. Ed. New Delhi: Vikas Publishing House PVT Ltd.
- Markey and Demeyer A.N.. 1986. "Dewey Decimal Classification Online Project : Evaluation of a Library Schedule and Index Integrated into the Subject Searching

- Capabilities of an Online Catalog.” Dublin, Ohio : OCLC, Office of Research. 최재황. 1988. *인터넷 학술 정보자원의 디렉토리 서비스 설계에 있어서 DDC 분류체계의 활용에 관한 연구*. 『정보관리학회지』 제 15권 제2 호. P.40에서 재인용.
- Oxford English Dictionary. 2nd ed.* Oxford: Clarendon Press, 1989. v.3
- Svenonius, E. 1983. “Use of Classification in Online Retrieval.” *Library Resources & Technical Services*, 27(1): 76-80
- Traugott. 1997. “The Role of Classification Scheme in Internet Resource Description and Discovery.” <<http://www.ub2.lu.se/desire/rade/r/reports/D3.2.3>>
- Vizine-Goetz, D. 1996. “Using Library Classification Schemes for Internet Resources.” *Proceedings of the OCLC Internet Cataloging Colloquium*. <<http://www.oclc.org/olcl/man/colloq/v-g.htm>>
- Buchanan, Brian 저. 정필모, 오동근 공역. 1989. 『문헌분류이론』. 서울: 구미무역.