

초등학생용 주제별 검색을 위한 효율적인 카테고리 분류 방법

정부현⁰, 김갑수
남양주창현초등학교⁰ 서울교육대학교 컴퓨터교육과
eduna@eduna.net kskim@snue.ac.kr

A Study on the Classification Scheme for the Design of Directory Kids Search Engines

Boo-Hyun Jeong⁰, Kap-Su Kim
ChangHyun Elementary School Dept. of Computer Education, Seoul National University of Education

요 약

인터넷을 통해 전달되는 교육자료의 양은 나날이 증가하고 있지만 정보 생산자들이 일정한 표준 없이 자의적인 기준에 의해 정보를 분류하여 구성하기 때문에 이용자가 필요한 정보를 정확하게 찾아내기란 매우 어려운 실정이다. 따라서 본 연구는 국내 주제별 검색엔진인 Yahoo Korea와 Naver, Hanmir, Empas의 초등학생용 검색엔진의 분류체계를 비교·분석하여 주제별 검색을 위한 효율적인 카테고리 분류 방법을 제시함으로써 정보접근에 익숙하지 않은 초등학생에게 쉽게, 빠르게, 정확하게 교육자료에 접근할 수 있는 분류체계를 제시하고자 한다.

1. 서 론

현재 방대한 인터넷 정보자원을 체계적으로 조직하여 이용자들이 필요로 하는 정보를 쉽고 빠르게 찾을 수 있도록 도와주는 다양한 웹 검색엔진들이 개발되고 있다. 하지만 이들 검색엔진들은 자체적인 주제분류체계와 다양한 성능의 탐색기능을 제공하고 있지만 아직 체계적으로 정보를 조직하지 못하고 있는 실정이다. 정보는 체계적으로 조직화되어 있을 때 그 가치가 커지게 된다. 초등학생들을 위한 체계적으로 조직화된 정보의 제공은 학생들에게 좋은 길잡이가 될 수 있으며 학교 수업 시간에도 충분히 이용할 수 있을 것이다.

인터넷을 통해 전달되는 교육자료의 양은 나날이 증가하고 있지만 정보 생산자들이 일정한 표준 없이 임의의 기준에 의해 정보를 분류하여 구성하기 때문에 이용자가 필요한 정보를 정확하게 찾아내기는 매우 불편한 것이 사실이다. 현재 방대한 인터넷 정보자원을 체계적으로 조직하여 이용자들이 필요로 하는 정보를 쉽고 빠르게 찾을 수 있도록 도와주는

다양한 웹 검색엔진들이 개발되어 탐색기능을 제공하고 있지만 이들은 아직 체계적으로 정보를 조직하지 못하고 있다. 많은 디렉토리형 검색엔진들의 주제 분류체계는 분류이론에 근거한 것이 아니라 임의의 계층구조를 따르는 경우가 대부분이어서 초등학생 이용자들은 탐색시간에 비해 만족할만한 탐색결과를 얻지 못하는 것으로 알려져 있다[1].

최근에는 디렉토리형 검색엔진의 분류체계에 이론적 배경을 도입하여 인터넷 상에 흩어져있는 수많은 자원을 체계적으로 분류함으로써 주제간의 상·하위 관계를 파악하고 관련 주제를 함께 모아 이용자의 탐색 노력을 줄이고 원하는 정보를 찾을 수 있는 바람직한 방법으로 여겨진다.

따라서 본 연구는 인터넷을 통한 정보접근에 익숙하지 않은 이용자 입장에서 초등학생용 주제별 검색엔진을 위한 카테고리 분류체계를 조사하여 분석함으로써 합리적이고 체계적인 카테고리 구성을 위한 모형구축을 시도하였다.

2. 선행연구

인터넷상의 온라인 정보검색은 정형화, 체계화되지 않은 정보의 특정상 효율성 높은 검색시스템이 필요하다. 이러한 필요성에 의해 현재 온라인 검색시스템에 체계적이고 논리적인 분류체계 적용이 검색 효율을 높일 것이라는 추정하에 많은 연구가 진행되고 있음을 볼 수 있다.

먼저 1983년에 온라인 검색에서 분류체계 적용이 검색 효율성(신속성, 정확성, 경제성)을 증진시킨다는 유용하다는 연구가 Svenonius(1983)에 의해 이루어졌다[2]. 이어서 Markey와 Demeyer(1986)는 DDC 온라인 프로젝트에서 이용자 주제접근을 개선시키기 위한 분류체계의 구현방안을 최초로 제시하였다. 1995년에 수행된 연구에서 Dahlberg(1995)는 네트워크 환경에서 LC와 DDC를 적용해서 분류이론 적용에 대한 적합성 여부를 분석하였는데, 각 주제의 분석과 추적 및 계층 구분에 유용함을 밝혀냈다[3].

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[4].

Wallice와 Burden은 웹자료의 대한 분류체계의 기반탐색을 분석하였다. DDC를 이용한 웹도서관과 같은 분류체계 기반 검색엔진의 성능을 분석하였다.

국내에서도 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[5].

1998년에는 최희윤(1998)이 인터넷 정보자료의 분류체계를 DDC와 비교분석하여 분류체계 설계 시 고려할 사항을 언급하였다. 또한 같은 해 남영준(1998)도 기존의 웹 문서 분류체계에 대한 비교·분석을 통해 웹 검색엔진

용 분류체계를 새롭게 설계하여 제시하였다. 최재황(1998)은 DDC 분류체계를 활용하여 문헌정보학분야 인터넷 학술정보자원에 대한 분류코드를 설계하였다.

신동민(2001)은 도서관 분류체계를 이용하지 않고 기존의 분류체계를 분석하여 현재의 디렉토리 검색엔진에서 사용하고 있는 틀과 비교분석하여 효과적인 디렉토리 구축을 위한 방안을 제시하였다[6].

한편 교육분야에서 이루어진 연구로는 이용우와 최성중(1999)의 “교육정보 체계화 및 표준화를 위한 다차원적 분류법”을 들 수 있다. 이 연구에서는 교육정보의 효율적이고 체계적인 분류 및 검색을 위해 분류체계의 7차원적인 접근방법을 제시하였다. 교육정보에 대한 내용, 지역, 시대, 대상, 매체, 교육방법, 용도에 의하여 교육정보를 다양한 분류체계로 분류하려 하였다.

이러한 연구들을 크게 둘로 구분해 본다면 하나는 일반적인 인터넷 자료들에 대한 연구로 앞에서 열거된 김영보(1997), 최재황(1998), 신동민(2001)의 연구가 여기에 해당한다. 또 다른 하나의 연구방향은 특정주제나 특정분야에 관련된 연구들이다. 전자책 콘텐츠 분야, 인터넷 쇼핑몰 상품, 산업분야, 정보통신기술 분야 등의 전문분야별 분류체계에 대한 연구가 이루어지고 있다[7].

최근에 이루어지는 연구들일수록 후자, 즉 특정주제나 특정분야에 관련된 연구가 많아지는 경향을 보이고 있다. 이러한 경향은 인터넷에서 제공되고 있는 검색엔진들은 모든 분야의 정보를 수집하여 분류하기 때문에 전문적인 검색을 원하는 이용자에게 유용한 정보원의 역할을 수행하지 못하므로 특정분야의 효율적인 정보검색을 위한 체계적인 디렉토리의 구성과 이용자 편이를 도모해야 할 필요성이 있기 때문이다.

3. 연구방법 및 범위

연구방법으로는 현행 국내 초등학교용 디렉토리 검색엔진들의 카테고리 분류체계를 조사·분석하고, 주제별 카테고리 분류의 원리와 계층구조를 밝히고자 한다. 구체적으로 국내 디렉토리 검색엔진 중에서 Yahoo Korea의 야후 꾸러기, Naver의 유니어네이버, Hanmir의 한미르개구쟁이와 Empas의 엠파스어린이를 선정하여 이들의 주제별 카테고리 분류체계를 비교하여 분석하고자 한다.

웹 디렉토리 검색엔진 분류체계의 하위구분은 계속적으로 세분화 및 통·폐합 등 가변적이기 때문에 분류체계를 고정시키기가 어려우므로 중구분 이하 하위구분은 조사 대상에서 제외시켰다.

검색엔진들의 계층관계에 있어서, 우선 최상위 계층(대구분 항목)과 하위계층(중구분 항목)의 주제어 수를 비교하고, 다음으로 최상위 계층의 주제어 유형을 비교한다. 그리고 동일한 명칭의 최상위 주제의 하위계층을 표본으로 추출하여 비교하고자 한다.

국내 초등학교용 주제별 검색엔진들의 카테고리 분류의 비교 및 분석된 내용을 고려하여 효율적이고 체계적인 분류체계의 모형을 구축할 것이나 인터넷 정보의 유동적인 특성에 따라 하위계층의 주제는 링크된 문서 수와 상위계층과의 논리적이고 일관성을 적절히 파악하여 적절하게 분류체계를 조절하고자 한다.

이 논문에서 구체적으로 다루고자 하는 내용을 순서대로 열거하면 다음과 같다. 첫째, 기존의 인터넷 검색엔진들이 제공하는 초등학교용 검색엔진의 카테고리 분류체계를 비교·분석한다. 둘째, 합리적이고 체계적인 카테고리 구성을 위한 모형구축을 시도하고자 한다.

4. 카테고리 분류 특성

4.1 야후 꾸러기의 카테고리 분류체계

야후! 꾸러기(<http://kr.kids.yahoo.com>)는 2000년 3월에 야후 코리아 내의 어린이 포털

서비스로 출발하여 어린이들에게 안전하고 유익한 디렉토리 구성과 다양한 서비스를 어린이들에게 제공하고 있다.

어린이들이 직접 인터넷 사이트를 쉽고 재미있게 검색할 수 있도록 만들어진 웹 도우미로서 교육적이고 어린이들에게 유익한 내용을 가진 사이트들을 전문 서퍼들이 직접 각 사이트를 방문하여 관련 카테고리를 만들었다.

카테고리 분류는 <공부와 학교>, <세계와 사회>, <자연과 과학>, <연예, 오락, 예술>, <운동, 놀이, 취미>, <컴퓨터, 게임> 총 6개의 상위 카테고리를 제공하고 있으며 그와 관련된 다수의 하위 카테고리들로 연결되어 있다.

<공부와 학교>라는 상위카테고리에는 1학년, 2학년, 3학년 등의 학년별, 국어, 수학, 사회 등의 과목별의 하위 카테고리들로 구성되어 있으며 카테고리 옆에 @ 표시는 다른 경로의 카테고리가 있음을 나타내었다. 즉 현재 페이지 이외에 다른 페이지에서도 똑같은 내용의 카테고리를 찾을 수 있음을 의미한다. 디렉토리 서비스 외에도 어린이들에게 추천하는 재미있는 게임, 동화, 동요, 학습 서비스 등 풍부한 정보를 제공하고 있다.

특히 야후는 분류항목의 반복성과 순환성을 특징으로 갖고 있는데, 반복성은 이마 하나의 체계에 속해 있는 주제어라 하더라도 필요에 따라서는 몇 번이고 반복해서 다른 체계에 속하게 되는 성질을 의미한다. 이는 DDC에서 형식구분을 두어 해당되는 분류번호에 따라 계속 사용할 수 있는 것과 유사한 면을 가지고 있다. 반복성을 설명하기 위해 <교육>이라고 하는 하나의 분류항목을 예로 들면, 초기에는 중분류 수준에 있으나 필요에 따라 하위 분류 수준에서도 출현한다. 순환성은 해당되는 주제 항목이 상위 수준에 포함되지 않았다 하더라도 @기호를 활용하여 다른 분류항목과의 링크를 통해 해당 정보를 입수할 수 있도록 하는 성질이다. 시스템 측면에서는 이러한 @를 통하여 동일한 주제를 여러 개의 분류항목에 넣어두는 낭비를 막고, 이용자에게는 다양한 접근점을 주어 찾기 쉽게 하는 성질을 가지

고 있다.

야후 꾸러기의 카테고리 분류체계는 <공부와 학교>, <세계와 사회>, <연예 오락 예술> 등 6개의 대구분 항목하에 185개의 중구분 항목의 분류체계를 가지고 있다. 6개의 대구분 항목에는 다양한 하위영역의 주제가 포함되어 있으며, 이는 실제적으로 이 검색엔진에 링크된 웹사이트의 수와 이용자의 탐색빈도 등이 고려된 실용적인 분류체계이다.

야후 꾸러기의 분류체계는 Yahoo!igans(<http://yahoolignas.yahoo.com/>)의 번역본으로서 우리 실정에 맞지 않는 하위 항목들을 제외되고 어린이들에게 꼭 필요한 항목들이 추가되었다. 예를 들면 science & nature 항목의 <science & nature picture>, <science & nature videos>, <science & nature sounds>등이 제외되고 <3학년>부터 <6학년>까지의 학년별 하위 주제 영역이 추가되었다.

<표 1> 야후 꾸러기의 카테고리 분류(2004. 06)

대구분 항목	중구분 항목		
공부와 학교	국어 수학 사회 과학 도덕 미술	영어 음악 체육 바른생활 즐거운생활 슬기로운생활	1학년 2학년 3학년 4학년 5학년 6학년
	겨울방학 과목별 교육방송 기초학력평가 단체, 클럽 대안교육 독서 마인드맵 성교육 숙제도우미	신문활용교육 여름방학 영재교육 온라인교육 유학, 어학연수 참고자료 창의성교육 체험학습 특수교육 학교	학교생활 학습지 행사, 대회 동화 방학숙제 사진 상담 직업, 진로 컴퓨터
세계와 사회	개인홈페이지 경제 교통 국가상징 국경일, 기념일 대한민국 문화 상담	선생님 세계 여러나라 세계 지역별 신화, 전설 어린이, 가정 언어 여행, 관광 역사	유아 음식과 요리 자선, 봉사 정부 종교 지리, 지도 직업, 진로
연예 오락 예술	귀신, 공포 그림, 미술 놀이 놀이공원, 유원지 뉴스와 신문 독서 라디오 만화, 애니메이션	문화 예술 어린이신문 연예인, 스타 영화 예술 온라인카드 운세 유머, 재미	음식과 요리 음악 장난감 채팅 캐릭터 텔레비전, 방송 패션, 미용 플레이

4.2 주니어네이버의 카테고리 분류체계

주니어네이버는 네이버에서 운영하는 어린이와 청소년을 위한 교육용 검색엔진으로써 주제별로 디렉토리 서비스도 제공한다. 99년 4월에 오픈한 주니어네이버(www.jrnaver.com)는 어린이 회원 180만명의 회원으로 가입되어 있는 국내 최초의 어린이 전용 인터넷 포털사이트이다. 어린이들에게 무해한 인터넷 사이트만으로 구성된 어린이 전용 디렉토리 서비스를 선보여 주목을 받고있다. 특히 이 디렉토리에 등록된 사이트는 검색 로봇을 이용해 자동으로 찾아낸 것이 아니라 네이버의 전문 서퍼들의 확인을 거친 후 등록을 하여 유해환경으로부터 분리하고 있다. 또한 인터넷 윤리시간, 신고본부 등을 운영해 어린이들에게 올바른 인터넷 사용법을 자연스럽게 익히도록 하고 있다.

이밖에도 무료 홈페이지, 웹 메일, 메신저 등 다양한 어린이 커뮤니티 서비스, 숙제도우미 등 다양한 학습 콘텐츠와 <주니어 게임>, <우리가 난다>와 같은 엔터테인먼트 서비스를 통해 어린이의 눈높이에 맞춘 유익하고 다채로운 내용을 갖추고 있다. 주니어네이버의 가장 큰 특징은 어린이 회원들이 직접 참여하고, 만드는 서비스가 많다는 것이다. 개설된 각종 어린이 클럽은 홈페이지 강좌, 게임, 취미, 연예, 팬클럽 등 다양하게 구성되어있다. 자신의 클럽을 쉽게 개설하여 클럽의 운영자가 되어 직접 운영할 수 있도록 지원하고 있다.

카테고리 분류체계를 살펴보면 <컴퓨터, 게임>, <숙제, 공부>, <연예, 오락, 예술>, <세계, 사회, 탐구>, <취미, 놀이, 스포츠>, <키즈 생활 정보> 6개 대구분 항목이 있으며 238개의 중구분 항목을 가지고 있다. 특히 도서관, 어린이 상담, 온라인 학습 등 하위 분류를 포함하는 <키즈 생활정보>를 두고 있다.

<표 2> 주니어네이버의 카테고리 분류(2004.06)

대구분 항목	중구분 항목		
컴퓨터 게임	게임랜드 게임동치2 게임이름 고전게임 공포게임 무미기 대포게임 모바일게임 별난게임 보드게임 색칠공부	심리테스트 웃입히기 오리게임 인기게임 재미있는 게임 캐릭터게임 컴퓨터, 비디오 게임 타자연습 테트리스 플래시게임 학습게임	검색엔진, 포탈 게임 다운로드 게임 대회 게임 잡지 길드 뉴스, 잡지 아바타 어린이 인터넷 온라인 예절 온라인 커뮤니티 이메일 보내기
	인터넷유행 인터넷 안전 자료 다운로드 슈니버활동하기 슈니버 이벤트 주니어네이버	채팅하기 카드 보내기 프로게이머 시험, 자격증 아래아 한글 엑셀 웹에디터	컴퓨터 용어사전 컴퓨터 초보자 파워포인트 포토샵, 페인트샵 프로그램 배우기 플래쉬 홈페이지 만들기
연예 오락 예술	가수 기획사 라디오 스타 연기학원 영화 문세	유머, 우스개 인터넷 방송 잡지 텔레비전 패션, 미용 SF, 공포	만화가 만화영화, 애니메이션 만화잡지 만화주제가 만화책 만화, 동호회, 단체
	연재만화 자작 캐릭터 캐릭터 플래시애니메이션 TV 만화	공연예술 극장, 극단 기관, 단체 미술관 시각예술	음악 이달의 공연정보 전시회 조형예술 축제, 행사

4.3 한미르 개구쟁이의 카테고리 분류체계

Hanmir는 한국통신 멀티미디어 연구소에서 개발한 키워드형 검색엔진으로 검색속도가 빠를 뿐만 아니라 유즈넷 검색, 일간지 기사검색, 일본 웹사이트 검색, 이미지와 사운드파일 검색 등을 제공하고 있다. 신문기사 검색과 일본 웹사이트 검색은 다른 검색엔진을 이용한 메타검색으로서 Hanmir 자체가 해당정보를 데이터베이스로 보유하고 있는 것이 아니다. 한미르개구쟁이는 한미르에서 운영하는 초등학생 대상의 어린이 전용사이트로 '어린이 인터넷 길잡이'를 목표로 내걸고, 어린이들이 유해한 환경에서 벗어나 안전하고 유익한 정보를 접할 수 있도록 안내하며 인터넷 이용, 학습, 놀이 등에 관한 다양한 서비스와 어린이들이 직접 참여해볼 수 있는 공간을 제공하고 있다. 메일 보내기와 홈페이지 만들기, 온라인 카드 보내기를 비롯해서 인기있는 스타들의 프로필과 소식을 들을 수 있는 <우리들의 스타>와 키우고 있는 애완동물을 사랑하고 추천

도 받을 수 있는 <애완동물>, 주어진 주제에 대해 내 생각을 마음껏 펼치고 다른 주제를 제안해볼 수도 있는 <토론해요>, 숙제할 때 도움을 주는 <숙제도우미>, 추천사이트와 인기검색어, 게임, 채팅, 클럽 등 개구쟁이 아지트 코너를 두고 있다.

한미르개구쟁이는 6개의 대구분 주제항목을 가지고 있으며, <인터넷,게임>, <학교,숙제,공부>, <연예,오락,예술>, <학부모,선생님>, <취미,놀이,운동>, <세계,자연,과학>항목이 상위분류에 속해 있다. 188개의 중구분 항목을 포함하고 있으며 특히 <학부모,선생님>주제를 상위주제로 설정하여 교사와 학부모의 가이드 및 상담 등 어린이뿐만 아니라 교사 및 학부모가 참여할 수 있는 장을 마련하고 있다.

<표 3> 한미르 개구쟁이의 카테고리 분류(2004.06)

대구분 항목	중구분 항목		
인터넷 게임	게임 검색엔진 게임어름으로 찾기 고전게임 길드, 동호회 모바일게임 아이큐게임 컴퓨터, 비디오게임 플래시게임	게임 자료실 게임정보 공포게임 무미기게임 보드게임 (41) 온라인게임 키우기게임 학습게임	게임 잡지 게임제작회사 그림, 이미지 대회, 이벤트 색칠공부 온라인쇼핑 프로게이머
	PC통신 온라인카드 인터넷-안전 자격증 채팅 컴퓨터학원 하드웨어	검색엔진 이메일(E-mail) 인터넷-예절 자료실 초보자-가이드 포탈사이트 홈페이지-만들기	소프트웨어 인터넷-가이드 인터넷유행 잡지, 웹진 커뮤니티 프로그램-배우기
학교 공부	국어 바른생활 슬기로운생활 음악 체육	도덕 사회 실과 자연, 과학 한자, 한문	미술 수학 영어 즐거운생활
	방학 온라인-학습 학습게임	성교육 진로, 직업 행사, 대회	역사 참고자료 현장체험학습
연예 오락 예술	SF, 판타지 놀이공원 연예인 오디션, 기획사 유머, 재미 인터넷방송 팬픽, 팬아트	공연, 행사 댄스, 합창 연예정보 우정, 사랑 음악 텔레비전 핸드폰, PCS	귀신, 공포 라디오 영화, 비디오 문세 인터넷노래방 패션, 미용

4.4 엠파스 어린이의 카테고리 분류체계

엠파스어린이(<http://kids.empas.com/>)의 검색엔진의 특징을 살펴보면 문장검색엔진으로써 자연어의 검색이 가능하며 검색 서비스와 디렉토리 검색이 가능하다. 또한 카테고리별

전문 서퍼를 두어 전문화된 정보 제공하고 있다.

웹사이트를 운영하는 기관 이름으로 해당 홈페이지를 찾아주는 기능(사이트 이름 검색)과 웹페이지가 없어진 경우라도 검색엔진이 수집해 놓은 텍스트 자료를 확인할 수 있는 미리보기 기능을 제공하고 있다. MP3 파일, 사운드 파일, 이미지 파일, 동영상 파일, 압축 파일 검색을 제공하고 있다.

한편 어린이들의 궁금증을 쉽게 검색할 수 있는 '검색박사'와 700여 종류의 만화 노래 등을 이용할 수 있는 플래시 나라, 300여 가지의 게임을 즐길 수 있는 '게임짱' 등 다양한 콘텐츠와 함께 스팸메일 차단 기능을 갖추고 있다. '검색박사' 코너를 이용하면 어린이들에게 유익한 정보를 안전하게 찾아볼 수 있고 어린이들이 궁금해하는 점을 직접 물어보고 답변해 줄 수 있는 어린이 지식검색을 추가하였다. 이외에 유해한 사이트와 스팸메일로부터 어린이를 보호하기 위해 초등학교 선생님들로 이루어진 선생님 자문단을 구성, 인터넷 모니터링, 각종 설문조사 및 게시판 활동 참여, 학부모 상담 등 다양한 활동을 지원하고 있다. 엠파스 어린이는 어린이들의 눈높이에 맞춘 다양한 콘텐츠를 제공, 교육과 즐거움을 동시에 충족시키고 폭넓은 경험과 지식을 접할 수 있도록 도움을 주는 인터넷 에듀테인먼트 서비스로서 제공되고 있다.

엠파스어린이의 카테고리 분류 체계를 살펴보면 <게임,컴퓨터>, <공부,숙제>, <과학,자연>, <사회,세계>, <연예,예술,놀이>, <어린이 생활>의 6개 대구분 항목과 140개의 중구분 항목으로 구성되어 있다. 다른 검색엔진과 주제 구성이 비슷하며 표기와 범위에 있어 약간의 차이를 보이고 있으나 주제별 분류가 거의 일치하고 있음을 알 수 있다.

<표 4> 엠파스 어린이의 카테고리 분류(2004.06)

대구분 항목	중구분 항목		
게임, 컴퓨터	고전게임 롤플레이게임 비디오게임 시뮬레이션 게임 액션 게임 어드벤처게임 온라인게임 컴퓨터, PC게임 플래시게임 학습게임	게임다운로드 게임 소식, 잡지 모바일게임 프로게이머 엠키즈 온라인 예절 이메일 인터넷 인터넷 안전 인터넷유행	자료실 포털, 검색엔진 홈페이지 만들기 소프트웨어 온라인카드 자격증 컴퓨터 기초 컴퓨터 용어사전 컴퓨터의 역사 컴퓨터 잡지 하드웨어
공부, 숙제	1학년 2학년 3학년 4학년 5학년 6학년 과학 국어 도덕	사회 수학 실과 영어 음악 체육 한자, 한문 방학숙제 월별 숙제	교육자료 독서, 글짓기 심교육 숙제도우미 시험, 경시대회 신문활용교육 여름캠프, 체험학습 참고자료 학부모 가이드
자연, 과학	3학년 4학년 5학년 6학년 기계 날씨 동물 물리	발명 생물 식물 에너지 우주 화학 환경 자연, 지형	과학, 자연사박 물관 과학뉴스 과학상식 과학실험 과학자 과학행사 기관, 단체, 연구소

5. 카테고리 분류체계 분석

이미 선행연구에서 지적한 바와 같이 온라인 환경에서도 기존의 분류체계를 이용하는 것이 검색효율성을 높여준다는 연구결과가 있듯이, 분석대상 검색엔진들은 최상위 계층에서 각각 적용한 문헌분류체계 전체 또는 일부를 적용하고, 하위계층에서는 온라인 시스템상의 특성과 제한점을 고려하여 분류 항목 수를 조절하여 색인시스템의 비대함을 함축적으로 조정 제공하는 특성을 갖고 있다.

분석대상 검색엔진 모두가 같은 수의 대구분 항목을 가지고 있었다. 그러나 각각의 하위 구분 항목의 수는 Yahoo(185개), Naver(238개), Hanmir(188개), Empas(140개)의 순서로 나타나 각 검색엔진마다 하위 영역의 수는 상당히 차이가 있는 것으로 드러났다. 모든 검색엔진에서 공통으로 나타나는 대구분 항목은 <공부,숙제>, <컴퓨터,게임>, <연예,오락,예술>, <세계,사회>, <자연,과학>, <운동,놀이,취미>이었으며, 네 가지 검색엔진에 동시에 나타나는 대구분 항목은 <공부,숙제>, <컴퓨터

터,게임>, <연예,오락,예술>이며, 나머지 <세계,사회>, <자연,과학>, <운동,놀이,취미>은 부분 통·폐합되어 나타났다. Naver의 <키즈 생활정보>와 Hanmir의 <학부모,선생님>, Empas의 <어린이 생활>은 독립적인 새로운 분야로 삽입되어졌다. 즉 어린이의 일상적인 생활정보를 다루는 검색엔진이기 때문에 실용적인 주제를 대구분 항목으로 가지고 있는 특징을 보여주고 있다.

<표 5> 카테고리 분류체계의 대구분 항목

야후푸러기	유니어네이버	한미르개구쟁이	엠피스어린이
공부와 학교 (47)	컴퓨터, 게임 (56)	인터넷, 게임(43) 학교, 숙제, 공부 (23)	게임, 컴퓨터 (31)
세계와 사회 (23)	숙제, 공부(36)	연예, 오락, 예술 (35)	공부, 숙제 (27)
연예 오락 예술 (25)	연예, 오락, 예술 (26)	학부모, 선생님 (13)	과학, 자연(23)
운동 놀이 취미 (48)	취미, 사회, 탐구 (38)	취미, 놀이, 운동 (42)	사회, 세계(13)
자연과 과학 (27)	취미, 놀이, 스포 츠(37)	세계, 자연과학 (32)	연예, 예술, 놀 이(20)
컴퓨터 게임 (15)	키즈 생활정보 (45)		어린이 생활 (26)
계: 6개	계: 6개	계: 6개	계: 6개

분석대상 검색엔진 간에 최상위 계층에서 중복하여 출현하는 주제 중 <공부,숙제>, <컴퓨터,게임>, <연예,오락,예술>가 가장 많이 제공되는 주제로서, 완전하게 동일한 명칭과 유사한 명칭 또는 부분일치 등을 합하여 총 6개의 최상위 주제가 공통적으로 나타나 비교적 포괄성을 지닌 것으로 나타났다.

검색엔진들의 계층관계에 있어서, 우선 최상위 계층과 하위 계층의 주제어 수를 비교하고, 그 다음으로 최상위 계층의 주제어 유형을 비교한다. 그리고 동일한 명칭의 최상이 계층 주제의 하위 계층을 표본으로 추출하여 비교하고자 한다. 하위 계층의 주제어 분석으로, 앞에서 보는 바와 같이 하위 주제어는 140개로부터 238개로 평균 188개로 편차가 있으며 각 검색엔진들의 주제 구성이 다른 것으로 평가할 수 있겠다. 웹 상에서 새로운 정보를 분류하기 위해 최신 주제를 많이 포함함을 보여 주고 있다.

이와 같이 중구분 항목 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.

이러한 분석결과를 기초로 하여 디렉토리 검색엔진에서 최상위 주제와 하위주제 간의 포괄성과 논리적인 계층관계의 기준을 정한다면 이용자의 주제 접근상의 혼란을 방지할 수 있으리라 판단된다.

6. 카테고리 분류체계 설계

6.1 카테고리 구성 지침

현행 디렉토리 검색엔진들을 분석한 바와 같이 주제를 나타내는 주제어의 표기, 주제어 간의 관계 및 이용자 인터페이스 등에 있어 현격하게 다양한 형태를 취하고 있어서, 이용자가 동일한 내용을 검색하기 위해 여러 검색엔진을 탐색할 때 어려움을 겪고 있다. 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 분석결과를 토대로 모든 디렉토리 검색엔진들에서 공통으로 채용하여 카테고리 분류 기준의 지침을 제시함으로써, 이용자가 원하는 정보를 찾기 위한 최소한의 노력으로 최대의 만족을 갖도록 하고자 한다.

1) 분류체계의 조건에 관한 지침

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

첫째, 유연성으로서 최상위 주제를 제외한 하위 카테고리에서 새로운 주제의 삽입과 유사 주제의 통합 및 분리, 세분 등 웹 문서의 생성 및 소멸과 해당 주제의 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.

셋째, 주제어 범위의 명확성으로서 주제어로 선정된 용어가 나타내는 개념의 범위가 명확해야 한다. 특히 현행 디렉토리 검색엔진들의 주제범위가 서로 일치하지 않는 것은 주제어의 용어 개념이 불명확함에 따른 결과이다.

2) 주제의 계층관계에 관한 지침

검색엔진의 계층성은 동일한 계층의 주제어수에 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, 특정 주제를 중점적으로 제공하는 검색엔진은 해당 주제의 계층관계를 특정적으로 유지하도록 한다.

둘째, 상하위 계층간의 계층성 유지에 있어서 적정 주제 수를 유지하도록 한다. 즉 상위 계층과 포함관계에 있는 주제는 하위 계층의 주제 수를 고려하여 세분하거나 통합하여 적절 수를 유지해야 한다.

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

3) 주제어 선정, 배열에 관한 지침

주제어 형태는 검색엔진들에서 매우 큰 다양성을 보이고 있으므로 주제 접근에 혼란을 초래하는 요소이다.

첫째, 복수개념의 주제어를 단일 표본으로 선정할 때 가급적 복수개념을 포함하는 상위 주제를 선택하여 선정한다.

둘째, 최상위 주제어의 선정은 현재 이용되고 있는 검색엔진들에서 공통적으로 사용하고 있는 주제를 우선 선정하고, 그 외 검색엔진의

특성에 따라 약간의 주제어를 별도로 선정하여 제공한다.

셋째, 주제어는 검색결과에 따라 상·하위 개념 또는 관련 개념으로 검색의 확장, 제한을 할 수 있도록 주제어간 상호관계를 정의하고, 필요시 검색화면에서 상호관계를 볼 수 있도록 인터페이스를 제공한다.

6.2 카테고리 분류체계 모형설계

초등학생용 주제별 검색엔진들의 카테고리 분류체계를 비교·분석하고 위의 카테고리 구성지침에 따라 어린이 검색엔진의 분류체계의 모형을 구축하였다.

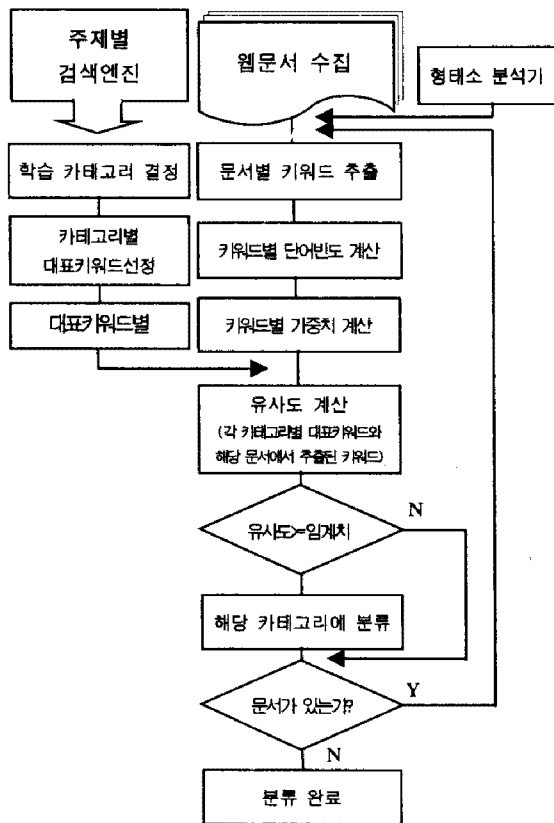
분류체계의 하위구분은 계속 세분화되는 추세를 보이고 있기 때문에 분류체계를 고정시키기가 어려우므로 중구분 이하 하위 구분은 설계에서 제외시켰다.

완성된 초등학생용 주제별 검색엔진의 카테고리 분류체계의 대구분 및 중구분 모형은 다음과 같다.

<표 7> 초등학생용 주제별 검색엔진의 카테고리 분류체계 모형

대구분 항목	중구분 항목			
공부와 학교	-국어 -수학 -사회 -과학	-도덕 -미술 -실과	-영어 -음악 -체육	-바른생활 -즐거운생활 -슬기로운생활
세계와 사회	-방학숙제 -독서 -성교육	-속재도우미 -신문활용교육	-온라인 교육 -참고자료 -체험학습	-행사, 대회 -진로, 직업
연예 오락 예술	-경제 -교통 -국경일, 기념일	-문화 -역사	-종교 -정치	-지리, 지도 -세계 여러나라
운동 취미	-뉴스와 신문 -방송 -애니메이션	-무용, 춤 -연예인, 스타 -영화	-예술 -유머 -음악	-캐릭터 -패션, 미용
자연과학	-축구 -태권도 -줄넘기 -스키	-스케이트 -농구 -배구 -수영	-씨름 -아구 -양궁 -육상	-베드민턴 -무용 -운동선수
컴퓨터 게임	-과학자 -날씨, 기후 -물리, 화학	-물질 -발명	-생물 -실험과 활동	-우주 -환경
	-온라인 예절 -이메일 -인터넷 -자료실	-홈페이지 제작 -소프트웨어 사전 -하드웨어	-컴퓨터 기초 -컴퓨터 용어 사전 -컴퓨터 잡지 -비디오게임	-PC게임 -롤러시게임 -학습게임 -게임소식, 잡지

카테고리 분류체계의 모형을 형성한 후 유입된 문서를 해당 주제에 맞게 분류하는 시스템을 설계하고자 한다.



<그림 1> 문서 수집 및 카테고리별 분류과정

<그림1>은 주제별로 분류된 문서를 수집하여 카테고리별로 분류하는 과정을 나타낸 것이다. 구체적으로 살펴보면 다음과 같다.

첫째, 카테고리 분류체계를 분석하고 필수 학습요소를 추출한다.

둘째, 필수학습요소를 토대로 학습카테고리를 먼저 선정하고, 학습카테고리와 관련된 필수학습요소, 웹문서, 컴퓨터용어사전 등을 이용하여 각 카테고리를 대표할 만한 키워드들을 결정한 후 가중치를 부여한다.

셋째, 형태소 분석기를 가지고 수집된 각 문서들로부터 키워드들을 추출한다.

넷째, 추출된 키워드들에 대한 단어빈도와 역문헌 빈도를 계산하고 이를 이용하여 키워드 가중치를 계산한다.

다섯째, 문서에서 추출된 키워들의 가중치와 학습 카테고리별 대표키워드들의 가중치를 이용하여 각 카테고리와의 유사도를 계산한다.

여섯째, 유사도에 의해 해당 카테고리에 문서를 분류한다.

7. 결론 및 제언

본 연구에서는 국내에서 사용되는 대표적인 초등학생용 검색엔진의 카테고리 분류체계를 조사하여 분석하고, 웹 상에서 사용될 수 있는 통일된 검색엔진의 카테고리 분류체계를 설계하였다. 분석 대상 검색엔진은 Yahoo Korea와 Naver, Hanmir, Empas의 초등학생용 주제별 검색엔지들이었다. 이들의 분류체계는 최상위 주제가 공통적으로 나타나 비교적 포괄성을 지닌 것으로 나타났지만 하위 분류 체계는 검색엔진 개발자들의 자의적인 판단에 따라 정한 것으로 볼 수 있으며 디렉토리의 계층성에 있어서 비논리적인 부분이 다소 나타났다. 대체로 각 검색엔진들이 다양한 형태의 분류체계 및 인터페이스를 유지함으로써 비능률성을 가지고 있었으며, 표준화된 디렉토리 검색엔진의 주제 분류체계에 대한 연구가 필요함을 확인하였다.

많은 양의 자료들을 초등학생 수준에 맞게 효율적으로 카테고리를 분류하여 제시한다면 사용자에게 큰 장점이 있을 것이다. 교육적 측면에서 초등학생들이 인터넷의 넓은 정보의 바다에서 시간을 허비하지 않도록 안내하기 위해 주제별 검색을 위한 효율적인 카테고리 분류 방법을 제시하였다. 체계화되지 않은 많은 교육자료들을 주제별 검색을 위한 효율적인 카테고리 분류 방법을 제시함으로써 정보 접근에 익숙하지 않은 사용자에게 쉽게, 빠르게, 정확하게 교육자료에 접근할 수 있는 분류체계의 표준화 가능성을 제시하였다.

분석결과로서 현행 디렉토리 검색엔진들은 주제선정에서의 논리성이 미약하고, 이용자에 대한 주제접근 편의성에 대한 배려가 부족하여, 각 검색엔진들이 다양한 형태의 인터페이스 및 문서 분류체계를 유지함으로써 디렉토리 검색엔진으로서의 카테고리 분류체계에 대한 개선 방안이 필요함을 확인하였다. 다만 인터넷상의 정보 유통은 그 양이나, 형태, 종류 등이 빠르고 다양하게 변화하고 있으므로, 계속적인 보완을 위해서 후속 연구 및 개정이 필요하다고 판단된다. 또한 현재 이용되고 있

는 많은 디렉토리 검색엔진들은 대부분이 전 주제 분야를 대상으로 하고 있는데, 향후 특정 주제 분야를 특성화하여 카테고리를 세분화하고 전문화하여 해당 주제분야의 분류체계 작성에 대한 지속적인 연구가 수행되어야 할 것으로 보여진다. 검색엔진의 개발시 주제의 전문성과 특정성을 고려하여 전문카테고리 분류체계의 참조가 필수적이라 하겠다. 이것을 위해서도 특정주제 분야의 카테고리 분류체계 작성에 대한 계속적 연구가 필요하다고 판단된다.

8. 참고문헌

- [1] Lancaster.FW, 저 장혜란 역 1998 색인초록의 이론과 실제, 구미무역
- [2] Svenonius. E, "Use of classification in online retrieval.", Library Resources & Technical Services, vol.27 no.1 pp.76-81
- [3] Dahberg, R, "Structuring and indexing the internet."
- [4] Vizine - Goetz. D. "Using library classification Schemes for internet resources." Proceedings of the OCLC Internet Cataloging Colloquium.
- [5] 김영보, "인터넷 탐색엔진의 분류체계에 관한 연구(:컴퓨터·인터넷 분야를 중심으로)", 석사학위논문, 성균관대, 1997.
- [6] 신동민, "인터넷 검색엔진의 디렉토리 구성에 관한 연구" 정보관리학회지 제 18권, 제 2호, pp.143-163, 2001.
- [7] 최재향, "인터넷 학술정보자원의 디렉토리 서비스 설계에 있어서 DDC 분류체계의 활용에 관한 연구", 정보관리학회지 제 15권, 제 2호, pp.47-68, 1998.