

# 인터넷 학술정보자원의 디렉토리 서비스 설계에 있어서 DDC 분류체계의 활용에 관한 연구

## A Study on the Use of DDC Scheme in Directory Search Engine for Research Information Resources on Internet

최재황(Jae-Hwang Choi)\*

### 목 차

- |   |                                 |
|---|---------------------------------|
| 1. 서론   | 4.1 시스템 구성도와 구성요소               |
| 2. 인터넷정보자원 관리와 DDC 분류체계                       | 4.2 인터넷정보자원 수집에 있어서 로봇 에이전트의 역할 |
| 3. 인터넷정보자원의 디렉토리 서비스에 있어서 DDC 분류체계의 활용 사례     | 4.3 작업 클라이언트의 내용                |
| 4. DDC 분류체계에 의한 문헌정보학분야 인터넷정보자원의 디렉토리 서비스의 설계 | 4.4 검색 클라이언트의 설계                |
|   | 5. 결론                           |

### 초 록

인터넷이 제공하는 학술정보자원은 풍부하지만 그 중에서 이용자들이 자신에게 필요한, 좋은 정보를 찾기는 쉽지 않다. 체계적으로 정리되어 있지 않기 때문이다. 본 연구에서는 도서관에서 오랫동안 이용해온 분류체계인 DDC(Dewey Decimal Classification)를 이용하여 학술분야 인터넷정보자원을 검색할 주제별 디렉토리 검색엔진을 설계하였다. 이 검색엔진을 설계함에 있어서 대개념에서 소개념으로 특정주제를 세분해가는 DDC의 '체계적 배열'에 따라 분류코드를 설계하였고, DDC의 '상관색인'을 이용하여 자동 문서분류사전을 작성하였다.

### ABSTRACT

Although the research information resources on Internet are spread out on thousands of computers, it is not always easy to get them on the right time by the right manner. The purpose of this study is to use DDC(Dewey Decimal Classification) scheme in subject-based directory search engine for research information resources to aid retrieval on the Internet. For the design of classification code, this study followed 'systematic order' of DDC to arrange subjects from the general to the specific in a logical order, and for the design of classification dictionary, 'Relative Index' of DDC was used to bring together the various aspects of subjects.

키워드 : DDC, 인터넷, 로봇 에이전트, 디렉토리 검색엔진

\* 성균관대학교 강사

■ 논문 접수일 : 1998년 7월 23일

## 1. 서론

인터넷은 1969년 미국 국방성 연구원들과 군 납업체, 관련 연구기관간에 정보공유를 실현하기 위한 허브기반 통신망의 구축과 함께 나타났으며, 인터넷의 멀티미디어 서비스라고 할 수 있는 WWW(World Wide Web)는 1989년 유럽의 CERN(European Particle Physics Laboratory)에서 각종 연구 자료와 정보들을 효율적으로 공유하기 위하여 만들어 졌다. 인터넷이 각종 정보의 수집 및 활용을 용이하게 하고자 출현하였고, 그 검색도구인 탐색엔진(search engine)들이 검색효율을 높여가고 있음에도 불구하고, 최종이용자들이 전세계적으로 약 3천 만개(1998년 1월 현재)(Network Wizards, 1998)의 호스트가 접속되어 있는 인터넷을 통하여 원하는 정보를 신속하게 얻는다는 것은 아직도 쉬운 일이 아니다. 인터넷의 정보자원은 매우 풍부하지만 체계적으로 잘 정리되어 있지 않기 때문이다.

인터넷에서 필요로 하는 좋은 정보를 얻기 위해서는 인터넷 정보의 여과 기능과 체계적인 조직이 중요하다. 현재 10여 개의 주요 탐색엔진들은 크고 작은 색인 데이터베이스와 다양한 탐색 기능을 제공하고 있지만 아직 체계적으로 정보를 조직하고 있지 못하며, 주제별 리스트를 제공하는 탐색 서비스에서도 사용되는 주제범주들 사이에 체계가 없다. 표준화된 주제 분류체계의 도입이 필요하다.

도서관에서 전통적으로 자료조직에 이용되는 DDC(Dewey Decimal Classification), LCC(Library of Congress Classification), UDC(Universal Decimal Classification)와 같은 분류체계를 인터넷정보자원 분류에 도입하는 경우는 드물다. Mitchell(1996)은 그의 최근 조사

에서 DDC, LCC, UDC의 적용 시도가 미약하나 몇몇 데이터베이스에 구조적, 내용적 변화를 가져오기 시작하였음을 지적하고 있다. 그에 의하면 DDC, LCC, UDC와 같은 분류체계가 비상업적 인터넷정보자원 접근에 가끔 이용되고 있으며 그런 예는 Morton Grove Online WEBrary(DDC), CyberStacks(sm)(LCC), NISS Information Gateway(UDC) 등에서 찾아 볼 수 있다고 하였다. 정보자료 정리의 전문가인 사서와 정보전문가들이 분류체계를 통한 인터넷 정보자원의 조직화에 관심을 가질 시기라고 본다.

그 동안 교육부 산하 첨단학술정보센터는 학문 각분야의 주제별 학술정보자원 구축을 목표로 '전문학술정보센터'의 지정과 육성을 위하여 많은 노력을 기울여 왔다. 이런 노력이 결실 할 때 최종이용자들이 인터넷에서 전문학술정보센터들을 통하여 정보를 쉽게 얻을 수 있도록 주제별 디렉토리서비스를 제공한다면 학술분야의 전문화, 정보화, 개방화라는 최근 조류에 상응하는 정보 유통체계의 토대를 마련하는데 크게 이바지 할 것으로 생각된다. 전문학술분야의 정보 수요가 급증하는 시점에서 인터넷상에 흩어져 있는 학술정보자원을 효과적으로 조직하는 일은 무엇보다 시급하고 중요하다고 본다. 인터넷상에 흩어져 있는 학술정보자원을 분야별로, 도서관에서 사용하는 분류체계를 이용하여 체계적으로 접근하도록 한다면 이용자의 탐색노력을 줄이고, 정보의 상·하위 관계를 보여주어 관련 주제를 함께 찾을 수 있게 될 것이다.

이런 뜻에서 본 연구는 첫째, 인터넷상에 흩어져 있는 한 학술분야, 즉 문헌정보학분야의 학술정보자원을 로봇 에이전트(robot agent)를 통하여 수집하고, 둘째, DDC 분류체계에 따라 이 분

야의 분류코드와 분류사전을 마련하고, 셋째, 문헌정보학분야의 인터넷정보자원을 자동 및 수동 분류한 후 디렉토리서비스를 설계하여 제시하고자 하는 것이다.

## 2. 인터넷정보자원 관리와 DDC 분류체계

문헌정보학자와 사서들은 도서관 분류체계가 정보에 대한 주제접근을 향상시킬 수 있다는 것을 오래전부터 인식해 왔다. Svenonius(1983)는 그의 논문에서 온라인 정보검색시스템에 있어서의 분류체계 이용에 대하여 이렇게 말하고 있다. 즉, 분류체계는 온라인 검색시스템에서 ① 재현률 또는 정확률을 향상시키고, ② 탐색용어에 문맥을 제공하고, ③ 브라우징을 가능케 하고, ④ 언어간 변환을 위한 수단에 독특한 기여를 한다는 것이다.

Markey와 Demeyer(1986)도 DDC 온라인 프로젝트에서, 최종이용자의 주제접근, 브라우징, 그리고 배열에 있어서 도서관분류체계의 탁월한 유용성을 역설한 바 있다. Markey와 Demeyer의 주요발견은 DDC의 계층자료(hierarchy data)로 주제접근이 향상되었고, DDC의 색인용어(index terminology)로 목록이용자들이 추가 관련항목을 검색할 수 있게 되었다는 것이다.

Svenonius와, Markey와 Demeyer에 의해 제기된 온라인 환경에서 분류체계의 도입 필요성은 즉각적인 반응을 일으키지 못하였다. 10여 년이 지난 후에야 OCLC(Online Computer Library Center)를 중심으로 인터넷정보자원에 대한 브라우징 구조 및 검색능력을 향상시킬 도구로서 DDC에 대한 연구에 관심이 일어나고 있다. 최근 연구자들은 DDC가 도서관자료와 인터

넷정보자원의 조직을 위하여 유용한 브라우징 도구가 될 수 있다는 점에 인식을 같이하고 있다(Vizine-Goetz, 1996). Koch(1997)는 분류체계에 의한 인터넷정보자원 조직의 장점을 다음과 같이 요약한다.

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): DDC는 계층구조로 되어 있으므로 필요시 탐색의 확대와 축소가 용이하다.

③ 문맥(context)제공: DDC 분류체계는 검색에 이용될 용어에 문맥을 제공한다. 예를 들어 동형어의어(철자구조가 같지만 뜻이 다른 단어)의 문제를 간단히 해결해 준다.

④ 다국어 접근의 가능성(Potential to Permit Multilingual Access): DDC에서는 숫자기호만을 사용하므로 다국어 자료 접근이 용이하다.

⑤ 분할 및 조작(Partitioning and Manipulation): DDC의 큰 주제항목은 필요시 논리적으로 작은 주제항목으로 세분되거나 조작될 수 있다.

⑥ 주지성(Potential to be well-known): 많은 도서관 이용자들은 전통적인 도서관분류체계인 DDC에 익숙해져 있다. 이들은 자기의 관심 주제 분야의 DDC 분류전개에 익숙해져 있을 가능성이 높다.

## 3. 인터넷정보자원의 디렉토리 서비스에 있어서 DDC 분류체계의 활용 사례

표 1의 WWW 사이트 명과 URL(Uniform

〈표 1〉 인터넷정보자원의 DDC 분류 사례

사이트 명	URL	성 격
1. biz/ed: Numeric List of BIZ-DEWEY-1 Sections	<a href="http://bized.ac.uk/roads/htdocs/subject-listing/Default/numlist.html">http://bized.ac.uk/roads/htdocs/subject-listing/Default/numlist.html</a>	biz/ed는 경영학 교육(business education)을 위한 인터넷 서비스다. 경영학 교육에 관심 있는 모든 사람을 위하여 자료들을 DDC로 분류하여 1996년부터 서비스하고 있다. DDC의 '330'과 '650'이 주류를 이루고 있다.
2. Blue Web'n Browse by Subject Area	<a href="http://www.kn.pacbell.com/wired/bluwebn/categories.html">http://www.kn.pacbell.com/wired/bluwebn/categories.html</a>	교육분야의 인터넷정보자원을 서비스하고 있다. 특기할 만한 것은 나열순서가 DDC분류순이 아니라 알파벳순이라는 것이다. DDC번호는 알파벳순 주제항목의 괄호 안에 표시되어 있다.
3. Browse LINK by DDC (BUBL)	<a href="http://link.bubl.ac.uk/ISC2">http://link.bubl.ac.uk/ISC2</a>	전체를 10개의 主類(000-900)로 나누고 主類 밑에 여러 개의 綱, 그리고 각 綱 밑에 目과 細目분류를 함께 하고 있다. 주제별(자모순) 브라우징도 가능하다.
4. Canadian Information By Subject (National Library of Canada)	<a href="http://www.nlc-bnc.ca/caninfo/esub.htm">http://www.nlc-bnc.ca/caninfo/esub.htm</a>	캐나다국립도서관(National Library of Canada)에 의하여 개발되었으며 캐나다에 관한 인터넷상의 모든 정보를 주제별로 제시하고 있다. 주제별(자모순) 브라우징도 가능하다.
5. CyberDewey	<a href="http://ivory.lm.com/~mundie/DDHC/CyberDewey.html">http://ivory.lm.com/~mundie/DDHC/CyberDewey.html</a>	전체를 10개의 主類(000-900)로 나누고 主類 밑에 다시 10개의 綱, 그리고 각 綱 아래 目 및 細目분류를 함께 하고 있다. DDC에 의해 모든 분류항목을 나열하고 있지만 빠진 분야도 꽤 많다. 주제별(자모순) 브라우징도 가능하다.
6. Expanding Universe: A Classified Search Tool for Amateur Astronomy	<a href="http://www.mtrl.toronto.on.ca/centres/bsd/astronomy/index.html">http://www.mtrl.toronto.on.ca/centres/bsd/astronomy/index.html</a>	천문학분야의 인터넷정보를 제공하고 있다. 천문학분야(DDC의 '520', '522', '523', '525')의 각 目 아래 細目분류를 하고 있다. Metropolitan Toronto Reference Library에 의하여 유지되고 있다.
7. GNOSIS: Global Network of Silicon Information Services	<a href="http://www.aimnet.com/~clancey/gnosis/">http://www.aimnet.com/~clancey/gnosis/</a>	전체를 10개의 主類(000-900)로 나누고 각 主類 아래 관련 주제들을 나열하고 있지만 자료가 빈약한 분야가 많다. 개인(Patrick W. Clancey)에 의하여 유지되고 있다.
8. Internet Public Library Online Texts Collection	<a href="http://www.ipl.org/reading/books/">http://www.ipl.org/reading/books/</a>	IPL(Internet Public Library) Online Texts Collection은 7,000개 이상의 인터넷정보를 저자(자모순), 표제(자모순), DDC의 세가지 방법으로 제시하고 있다. DDC분류는 전체를 10개의 主類(000-900)로 나누고 각 主類 아래 綱과 細目분류를 함께 하고 있으며 目 아래 細目분류도 하고 있다.

〈표 1〉 인터넷정보자원의 DDC 분류 사례(계속)

사이트 명	URL	성 격
9. The Internet Resource (Napier University)	<a href="http://www.napier.ac.uk/depts/library/intres/ir000999.html">http://www.napier.ac.uk/depts/library/intres/ir000999.html</a>	The Internet Resource는 Napier대학교도서관 직원들에 의하여 유지되고 있다. 전체를 10개의 主類(000-900)로 나누고 主類밑에 綱, 그리고 각 綱밑에 目분류를 하고 있다. CyberDewey와 마찬가지로 DDC에 의한 모든 분류항목을 나열하고 있지만 제시하고 있는 인터넷정보자원은 양적으로 빈약하고 目아래 해당 URL이 없는 경우도 많다.
10. Internet Resources in Dewey Decimal Order (Mid-Continent Public Library)	<a href="http://www.mcpl.lib.mo.us/dewey.htm">http://www.mcpl.lib.mo.us/dewey.htm</a>	전체를 10개의 主類(000-900)로 나누고 각 主類아래 綱과 目분류를 함께 하고 있다. 빠진 분야가 많고 자료도 양적으로 빈약하다.
11. Mr. Dui's Topic Finder	<a href="http://www.oclc.org/oclc/fp/mrdui/mrdui.htm">http://www.oclc.org/oclc/fp/mrdui/mrdui.htm</a>	OCLC는 인터넷정보자원의 최종이용자 탐색을 용이하게 하는 도구로서 DDC를 이용한 몇 개의 프로젝트를 수행하여 왔다. 그 중의 하나가 Mr. Dui's Topic Finder라고 불리는 브라우저의 개발이었다. Mr. Dui's Topic Finder는 DDC를 기반으로한 NetFirst의 탐색을 가능하게 해주고 있다. NetFirst는 1996년 9월 현재 55,000여 개의 레코드를 포함하고 있으며, 매주 갱신되고 있음은 물론, 매일 수천개의 레코드가 추가되고 있다.
12. Mt. Laurel Hartford School's Library Without Walls (Resources for Students)	<a href="http://www.voicenet.com/~srussell/student.html">http://www.voicenet.com/~srussell/student.html</a>	초등학교 5-6학년용 대상으로 숙제(assignments)에 도움을 줄 목적으로 서비스하고 있다. 전체를 10개의 主類(000-900)로 나누고 각각의 主類아래 숙제에 도움이 될 만한 자료들을 제시하고 있다.
13. PICK: Quality Internet Resources in Library and Information Science)	<a href="http://www.aber.ac.uk/~tplwww/e/contents.html">http://www.aber.ac.uk/~tplwww/e/contents.html</a>	University of Wales Aberystwyth의 Thomas Parry 도서관에 의하여 유지되고 있으며 문헌정보학분야 인터넷정보자원을 제시한다. 문헌정보학분야인 DDC의 '020-028' 외에 관련분야도 주제항목으로 선정하여 DDC분류 순으로 제시하고 있다.
14. Sites by DDC (Alastair G. Smith)	<a href="http://www.vuw.ac.nz/~agsmith/bkmark.htm#ddc">http://www.vuw.ac.nz/~agsmith/bkmark.htm#ddc</a>	뉴질랜드의 Victoria University of Wellington에 의하여 유지되는 인터넷정보자원 서비스다. 전체를 10개의 主類(000-900)로 나누어 모든 학문분야를 포함하고 있지만 대부분의 자료가 컴퓨터와 문헌정보학분야에 치중되어 있다.

〈표 1〉 인터넷정보자원의 DDC 분류 사례(계속)

사이트 명	URL	성 격
15. Subject List In Classification Order (University College Chester)	http://chandra.chester.ac.uk/~smilne/chesweb2.htm	전체를 10개의 主類(000-900)로 나누고 각 主類 아래 綱目, 細目분류를 함께 하고 있다. 빠진 분야가 많고 자료도 양적으로 빈약하다.
16. Webrary Links Menu (Morton Grove Public Library)	http://www.webrary.org/ref/weblinksmenu.html	전체를 10개의 主類(000-900)로 나누고 각 主類 아래 綱분류까지만 하고 있다. 각 항목에 간단한 주석을 달고 있다.
17. WWLib Browse Interface (Wolverhampton Web Library)	http://www.scit.wlv.ac.uk/wwlib/browse.html	전체를 10개의 主類(000-900)로 나누어 브라우징 서비스를 제공해주고 있지만 主類이하에서는 DDC 번호를 입력하여야 한다. 예를 들어 '020' 을 입력하면 '020' 에서부터 '029.9352' 까지 총 25개 항목의 인터넷정보를 찾아낼 수 있다. DDC 20판을 이용하고 있으며 자료가 양적인 면에서 빈약하다.

Resource Locator)들은 1998년 5월 30일 현재 DDC 분류체계로 인터넷정보를 분류·제공하고 있는 것들이다. 필자는 인터넷상에 존재하는 모든 관련 사이트를 찾아 열거하려고 노력하였으며, McKiernan(1997)의 조사가 이들 17개의 WWW 사이트를 찾아 표에 수록하는데 도움이 되었다.

이들 17개 WWW 사이트들을 요약해 보면, 13개가 지식의 쏠분야(DDC의 000-900)를 대상으로 하고 있고, 나머지 4개 사이트만이 특정 분야를 전문적으로 분류한 것이다. 즉, ① biz/ed는 경영학교육을, ② Blue Web'n Browse by Subject Area는 교육분야를, ⑥ Expanding Universe는 천문학분야의 인터넷정보자원을, 그리고 ⑬ PICK은 문헌정보학의 관심분야를 전문 주제항목으로 선정하여 관련 사이트를 연결시켜

주고 있다. ④의 Canadian Information by Subject는 DDC의 전분야를 포함하고 있지만 캐나다에 관한 인터넷상의 정보만을 제시하고 있다. 한편, 개인에 의해서 유지되는 것으로서 ⑤의 CyberDewey(David A. Mundie)와 ⑦의 GNOSIS(Patrick W. Clancey)가 있으며, 이들 사이트들은 대학이나 기관 혹은 도서관에 의하여 운영되는 경우에 비하여 자료가 양적으로 빈약하다.

#### 4. DDC 분류체계에 의한 문헌정보학 분야 인터넷정보자원의 디렉토리 서비스의 설계

##### 4.1 시스템 구성도와 구성요소

본 연구에서 설계된 문헌정보학분야 인터넷정보자원의 디렉토리 서비스를 위한 시스템의 구성도와 구성요소는 다음과 같다.

#### 4.1.1 시스템 구성도

시스템의 구성도는 그림 1과 같으며, 크게 다음과 같은 내용과 과정을 가진다.

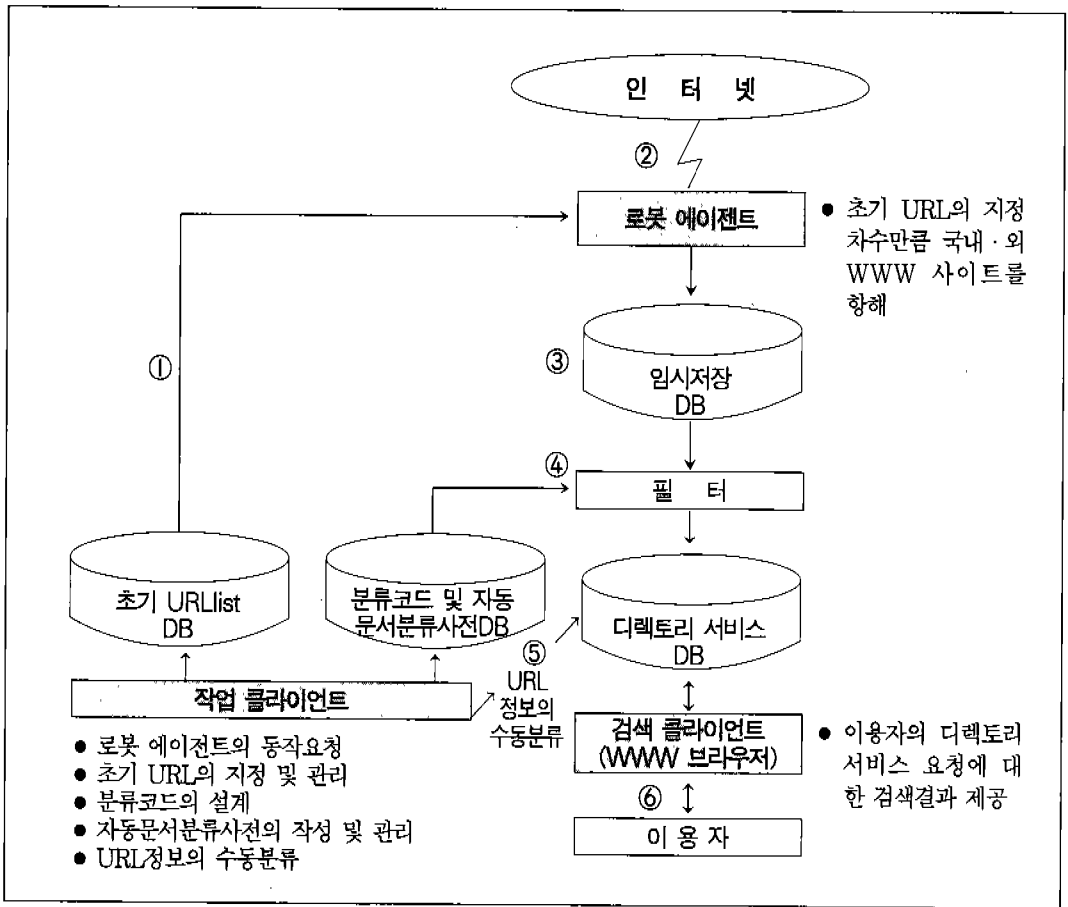
① 로봇 에이전트가 처음 작동 시 정보를 수집하기 위해서는 방문할 초기 URL이 필요하며, 이들은 '초기URLlist DB'에서 저장, 관리된다.

② 로봇 에이전트는 초기 URL로 서버에 질의 메시지를 보내고, 서버는 응답메시지를 보내온다.

③ 응답메시지는 '임시저장 DB'에 저장된다.

④ '필터'는 '분류코드 및 자동 문서분류사전 DB'를 참조(load)하여 '임시저장 DB'로부터 읽어들이는 개별 URL을 대상으로 필터링한 후 '디렉토리 서비스 DB'에 저장한다. 이때 작업은 모두 필터가 자동으로 처리한다.

⑤ '필터'에 의해 자동으로 분류 확정되지 않은 모든 데이터는 작업 클라이언트의 작업모듈에서 Dragdrop(마우스로 끌고 와서 집어넣는) 방



〈그림 1〉 시스템 구성도 개관

식 혹은 입력방식으로 수동 분류되거나 관련 없는 자료는 삭제된다.

⑥ 이용자는 검색 클라이언트(WWW 브라우저)를 이용하여 디렉토리서비스를 요청하고 검색 결과를 받는다.

### 4.1.2 시스템 구성요소

그림 1의 시스템 구성도는 다음과 같은 시스템 구성요소를 가지며, 크게 세 부분으로 나누어 볼 수 있다.

① 로봇 에이전트 모듈: 작업 클라이언트의 로봇 에이전트 동작 요청에 따라 초기 URL의 지정 차수(depth)만큼 국내·외 WWW 사이트를 향해 하며 정보를 수집한다(4.2 참고).

② 작업 클라이언트 모듈: 로봇 에이전트에 대한 모든 작업을 관리한다(4.3 참고).

로봇 에이전트의 동작 요청(4.3.1 참고)

초기URL의 지정 및 관리(4.3.2 참고)

분류코드의 설계(4.3.3의 a참고)

자동 문서분류사전의 작성 및 관리(4.3.3의 b참고)

URL정보의 수동분류(4.3.4 참고)

③ 검색 클라이언트 모듈: 디렉토리 서비스 요청에 대한 검색결과를 이용자에게 서비스한다(4.4 참고).

## 4.2 인터넷정보자원 수집에 있어서 로봇 에이전트의 역할

인터넷상에 널리 흩어져 있는 정보를 찾을 수 있도록 도와주는 검색엔진들은 정보의 수집이라는 측면에서 보면 다음의 두 가지로 나눌 수 있다(박정훈, 1995).

① 로봇 에이전트를 통하여 자료를 수집하는 검색엔진

② 다른 검색엔진이 보유하고 있는 데이터베이스를 이용하여 자료를 수집하는 검색엔진

로봇 에이전트를 통해 자료를 수집하는 검색엔진은 자료를 수집하는 로봇 에이전트, 수집된 자료를 저장하는 데이터베이스, 그리고 사용자가 질의를 했을 때 자료를 검색해주는 정보검색시스템으로 구성된다. 다른 검색엔진의 데이터베이스를 이용하는 검색엔진을 메타검색엔진이라고 하며, 이것은 전체적으로 질의를 각 검색엔진에 보내는 부분과 검색되어온 결과들을 통합하여 이용자에게 제공하는 부분으로 구성된다. 본 연구에서는 로봇 에이전트를 통하여 정보를 수집하는 검색엔진에 대하여만 논의한다.

### 4.2.1 로봇 에이전트의 기능

Robots, Spiders, Wanderers, Web Crawlers, Worms, 또는 Ants라고도 하는 WWW 로봇 에이전트는 인터넷상의 호스트들을 자동으로 방문하여 관리자를 대신해서 관리자가 원하는 작업을 자동적으로 수행하는 소프트웨어(프로그램)이다(최중민, 1997). 로봇 에이전트는 인터넷상에서 지정된 URL을 지정차수만큼 향해하며 HTML(Hypertext Markup Language)로 구축된 문서 정보를 얻어 데이터베이스에 저장하고 이용자로 하여금 해당분야에 대한 다양한 종류의 정보를 찾을 수 있게 한다.

로봇 에이전트는 통계분석(statistical analysis), 자원발견(resource discovering) 등의 목적으로 사용될 수 있지만(Koster, 1995), 다음과 같은 몇 가지 문제점을 가지고 있다.



① 로봇 에이전트의 표준이 없다는 것이다. 다양한 로봇 에이전트가 작성되고 데이터가 공유되려면 표준이 필수적이다. 그러나 현재까지 그러한 표준은 없고 접근에 대한 표준(3.2.4 참고)만 제안되었을 뿐이다.

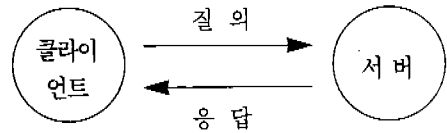
② 로봇 에이전트는 브라우저와 달리 이미지 맵, 폼 등을 처리하지 못한다. 로봇 에이전트 중 어떤 것은 이미지를 읽어 세밀하게 URL을 조정하는 능력이 있었으나 서버의 부담만 가중시키는 흠이 있었다. 즉, 이미지 맵을 사용하면 로봇 에이전트가 그 정보를 읽고 판단할 능력이 없다.

③ 로봇 에이전트는 ID와 암호가 필요한 서버에는 접근하지 못한다. 특정 서버만 접속하는 로봇이라면 큰 문제가 없겠지만 많은 웹서버에 필요한 ID를 모두 만든다는 것은 현재로서는 불가능하다.

#### 4.2.2 로봇 에이전트의 기본 알고리즘

로봇 에이전트는 원하는 정보를 얻기 위하여 HTML로 구축된 문서들을 순환적으로 검색하면서, WWW의 하이퍼텍스트 구조를 자동적으로 추적하여 정보를 저장해주는 프로그램이다. 이러한 문서들을 순환적으로 검색하기 위해서는 WWW의 프로토콜인 HTTP(Hypertext Transfer Protocol)를 준수하는 프로그래머이어야 한다. OSI(Open Systems Interconnection) 7 Layer에 의하면 HTTP는 Application 계층에 속하는 프로토콜로서, 클라이언트의 질의에 따라 서버가 해당되는 텍스트나 정보들을 클라이언트에게 응답하는 방식으로 되어 있다(김성훈, 1996)(그림 2 참고).

여기에서, 데이터의 처리나 전송을 요청하는 컴퓨터 처리기를 클라이언트(client)라 하고, 데이터를 처리해 주거나 전송해 주는 컴퓨터 처리



(그림 2) 로봇 에이전트의 기본 알고리즘

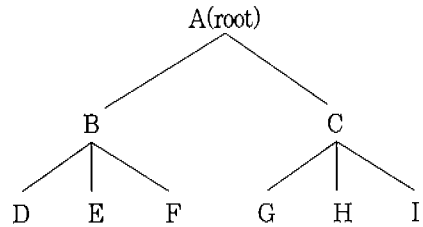
기를 서버(server)라 한다. 또한, WWW 서버와 클라이언트 사이에 하이퍼텍스트 문서를 전송하는 프로토콜로서 HTTP가 사용되며, 문서의 내부 형식을 규정하기 위한 표준 언어로서 HTML이 사용된다. 또한 인터넷상의 특정한 문서에 접근하는 방법과 문서의 소재 및 이름을 지정하는 수단으로서 URL이 사용되는 것이다(정영미, 1995).

#### 4.2.3 로봇 에이전트 프로그램의 수행방법

로봇 에이전트가 인터넷 호스트들을 대상으로 문서를 탐색하는 과정에는 넓이우선 탐색(breadth-first traversal)과 깊이우선 탐색(depth-first traversal)의 두 가지 방법이 있으며, Tree 구조로 설명될 수 있다(이영식, 황호정, 1996)(그림 3 참고). 본 연구에서는 하나의 호스트를 먼저 검색하고 다음 호스트를 검색하는 깊이우선 탐색방법을 적용하였다.

##### ① 넓이우선 탐색

이것은 한 다큐먼트에 여러 레퍼런스가 있을



(그림 3) URL로 본 Tree 구조

경우 root에 대하여 먼저 검색을 하고, 다음에 차례로 지정된 차수까지 검색하는 방법이다.

예: A → B → C → D → E → F → G → H → I  
(그림 3 참고)

② 깊이우선 탐색

이것은 한 다큐먼트에 여러 레퍼런스가 있을 경우, root에 대하여 먼저 검색을 하고, 다음에 어느 한 레퍼런스를 선택한 후 그것을 하위개념으로 계속 따라가면서 지정된 차수까지 검색하는 방법이다.

예: A → B → D → E → F → C → G → H → I  
(그림 3 참고)

4.2.4 로봇 배제의 표준

로봇 에이전트들은 우리가 인터넷을 항해하는 동안 바로 옆을 지나가는 것과 같은 것이므로 서버를 혼란스럽게 하는 일이 있다. 적합하지 않은 WWW서버의 부분들을 검색 한다든가, 일시적 가치밖에 없는 정보를 검색하는 일이 발생하게 된 것이다. 이러한 일들로 서버의 어느 특정 부분을 로봇들이 접근하지 못하도록 하는 방법을 세계의 로봇제작자들이 협의하여 제안하게 되었으며, 이것이 바로 '로봇 배제의 표준(A Standard for Robots Exclusion)'이다.

로봇 배제의 표준은 웹서버에 대한 로봇 에이전트의 접근 방법을 명시한 것으로, 로봇 에이전트의 너무 빠른 검색이나, 같은 파일의 계속 검색 등으로 인한 속도 저하를 방지하기 위한 것이다. 이 표준을 사용하려면 웹 서버에 '/robots.txt'를 마련하면 되고, 따라서 아무 내용도 적지 않은 '/robots.txt'는 어떤 로봇이 들어와도 환영한다는 의미가 된다. WebClawer의 Martijn Koster (1993, 1996)는 자신이 경험을 통하여 얻은 Robot 제작에 있어서 지켜야할 것들과 그 평가에

관한 내용을 담은 페이지를 만들어 놓기도 하였다. 본 연구에서도 이러한 약속의 규칙인 '로봇 배제'를 따르게 설계되었다.

4.3 작업 클라이언트의 내용

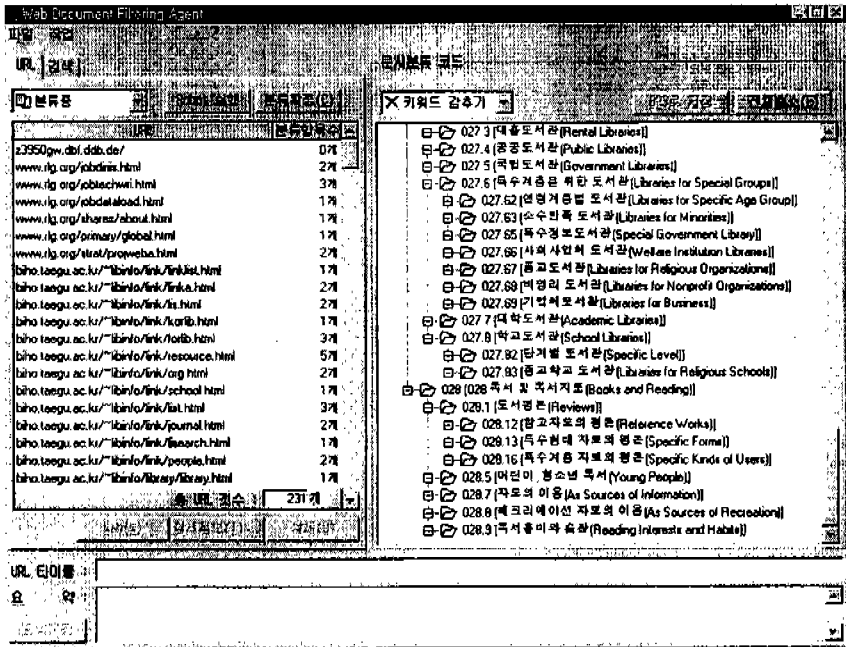
그림 4에 표시된 작업 클라이언트의 주 모듈은 로봇 에이전트가 수집한 인터넷정보자원을 관리자가 쉽게 관리할 수 있게 하는 모니터링 도구이다. 자동분류된 데이터로 URL 리스트를 만들고, 이를 편집할 수 있도록 하며, 현재 분류된 상태를 디스플레이 하게 된다. 이 내용을 아래에서 좀 더 자세히 설명한다.

4.3.1 로봇 에이전트의 동작 요청

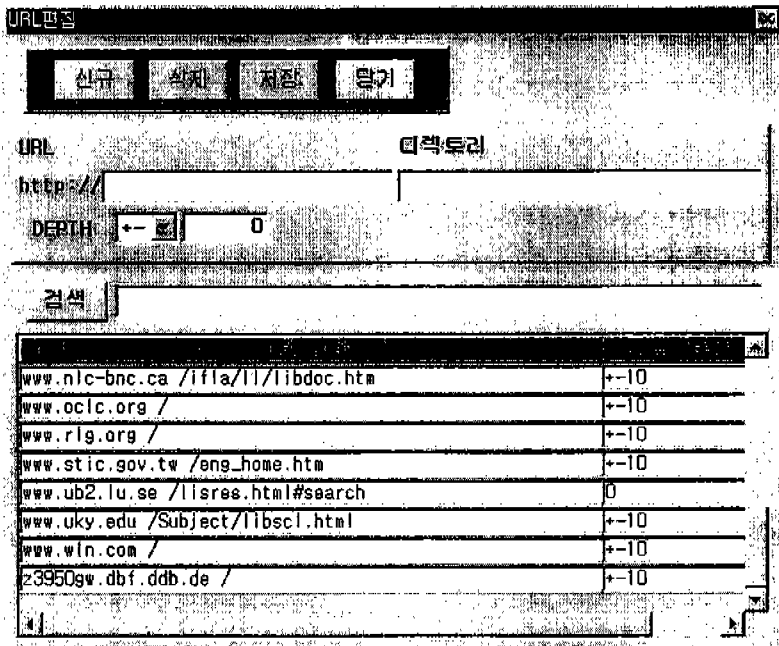
작업 클라이언트에서 초기 URL리스트를 입력받아 로봇 에이전트의 동작을 요청한다. 로봇 에이전트를 동작하는 방법에는 서버상에서 시간마다 자동 동작시키는 방법과 클라이언트에서 임의로 동작시키는 두 가지 방법 모두 가능하다.

4.3.2 초기 URL의 지정 및 관리

로봇 에이전트는 초기 URL정보에 따라 인터넷을 순회하도록 되어 있다. 그러므로 원하는 정보를 획득하기 위해서는 효율적인 초기 URL을 제시해 주어야 하며 각 초기 URL에 대해서는 차수까지 지정해 주어야 한다. 예를 들어, 차수를 '+-10'으로 정해주었다면 로봇 에이전트는 지정된 초기 URL에서 위로 10개, 아래로 10개의 레퍼런스까지 검색하게 되며, 차수를 '0'으로 지정해주었다면 지정된 초기 URL에서 위, 아래의 모든 레퍼런스를 검색하게 된다. 초기 URL을 지정해 주는 목적은 유용한 정보를 보다 많이 얻기 위한 것이므로 잘 알려진 자료보관 사이트를 중



〈그림 4〉 작업 클라이언트의 주 모듈



〈그림 5〉 초기 URL 관리 모듈

〈표 2〉 문헌정보학분야의 초기 URL 리스트

사이트 명	URL
1. OCLC (Online Computer Library Center)	<a href="http://www.oclc.org">http://www.oclc.org</a>
2. LC (Library of Congress)	<a href="http://www.loc.gov">http://www.loc.gov</a>
3. RLIN (Research Libraries Information Network)	<a href="http://www.rlg.org">http://www.rlg.org</a>
4. WLN (Western Library Network)	<a href="http://www.wln.com">http://www.wln.com</a>
5. NACSIS (The National Center for Science Information Systems)	<a href="http://www.nacsis.ac.jp">http://www.nacsis.ac.jp</a>
6. UTLAS (University of Toronto Libraries)	<a href="http://www.library.utoronto.ca/www/index.html">http://www.library.utoronto.ca/www/index.html</a>
7. BLAISE (The British Library's online information service)	<a href="http://blaiseweb.bl.uk">http://blaiseweb.bl.uk</a>
8. NISS (National Information Services and Systems)	<a href="http://www.niss.ac.uk">http://www.niss.ac.uk</a>
9. ABN (National Library of Australia)	<a href="http://www.nla.gov.au">http://www.nla.gov.au</a>
10. 문헌정보학 엘로우페이지	<a href="http://biho.taegu.ac.kr/~libinfo/link">http://biho.taegu.ac.kr/~libinfo/link</a>
11. LIBRARY AND INFORMATION SCIENCE, LIBRARIANSHIP	<a href="http://www.ub2.lu.se/lisres.html#search">http://www.ub2.lu.se/lisres.html#search</a>
12. IFLA LIBRARY AND INFORMATION SCIENCE: General Resources	<a href="http://www.nlc-bnc.ca/ifla/II/libdoc.htm">http://www.nlc-bnc.ca/ifla/II/libdoc.htm</a>
13. Library Information Resources on the Internet	<a href="http://www.library.nwu.edu/resources/library/libinfo.html">http://www.library.nwu.edu/resources/library/libinfo.html</a>
14. Internet Library for Librarians	<a href="http://www.itcompany.com/inforetriever/catinet1.htm">http://www.itcompany.com/inforetriever/catinet1.htm</a>
15. Library and Information Science Resources	<a href="http://lcweb.loc.gov/global/library">http://lcweb.loc.gov/global/library</a>
16. Library Resource List	<a href="http://www.dpi.state.wi.us/www/lib_res.html">http://www.dpi.state.wi.us/www/lib_res.html</a>

〈표 2〉 문헌정보학분야의 초기 URL 리스트(계속)

사이트 명	URL
17. Libraries & the WWW: Selected Resources	http://www.ualberta.ca/~nfriesen
PICK: quality Internet resources in 18. library and information science selected by Thomas Parry Library	http://www.aber.ac.uk/~tplwww/e
19. Innovative Internet Applications in LIBRARIES	http://pages.cthome.net/library/ innovate.html
20. Library & Related Resources	http://www.ex.ac.uk/~ijtilded/lib/ wwwlibs.html
21. World-Wide Web Resources - Library Science	http://www.uky.edu/Subject/libsci.html
22. Librarians' Index to the Internet: Libraries	http://sunsite.berkeley.edu/InternetIndex

심으로 조금씩 첨가해 가는 것이 바람직하다. 그림 5의 초기 URL 관리 모듈은 로봇 에이전트가 참조할 URL정보를 관리하는 화면을 나타낸 것이며, URL정보를 추가, 삭제, 편집하는 기능을 포함하고 있다. 본 연구에서 사용할 문헌정보학분야의 초기 URL은 표 2와 같다.

#### 4.3.3 분류코드의 설계와 자동 문서분류사전의 작성 및 관리

분류코드를 설계하고 자동 문서분류사전을 작성, 관리하는데 DDC 분류체계를 이용한다. DDC는 1876년에 처음 공刊된 이래, 현재 21판이 간행되고 있으며 세계에서 가장 널리 사용되는 분류표이다. 1997년 현재 135개 국가에서 사용되고 있으며, 30개 이상의 언어로 번역되고 있다. 문헌분류상에 DDC가 기여한 바는 두 가지 측면이 있다. 첫째는 단순한 숫자만을 사용하여

체계적인 순서로 주제를 배열하였다는 점이고, 둘째는 상관색인을 가짐으로써 어떤 주제와 관련된 분산된 관련 주제들을 알파벳순으로 한 곳에 모아줌으로써 체계적 순서에서 오는 이용상의 단점을 해결했다는 점이다. 여기서 체계적인 순서란 특정 주제나 형식 등을 대개념에서 소개념으로 점차 세분하여 간다는 뜻이다. 본 연구에서는 DDC의 문헌 분류 원칙인 주제의 체계적 배열 순서에 따라 분류코드를 설계하고, DDC의 상관색인을 이용하여 자동 문서분류사전을 작성한다.

##### (a) 분류코드의 설계

DDC는 십진식 분류법이며, 십진식 분류법이란 분류하고자하는 모든 대상을 0에서 9까지 10구분한다. 이처럼 DDC는 인류가 기록한 모든 지식을 10개의 主類(main classes)아래 모으고, 각 주류는 다시 10개의 綱(divisions)아래, 각 강은

다시 10의 목(sections)아래 모은다. 본 연구에서 문헌정보학 분야에 적용시킨 분류코드는 DDC의 주류, 강, 목을 채택하는 대신 대분류(주제범주), 중분류(1차 하위주제), 소분류(2차 하위주제)의 3단계를 채택한다. DDC에서 문헌정보학이 크게 위치하고 있는 부분이 '020'에서 '028'이므로 (DDC 21판) 주류부터 분류하는 것이 불가능하기 때문이다. 그러므로 실제로 분류한 단계는 綱, 目, 그리고 細目的 단계가 된다. 분류코드의 설계 결과는 표 3에 표시된 바와 같이 8개의 대분류, 39개의 중분류, 그리고 49개의 소분류가 나타났다.

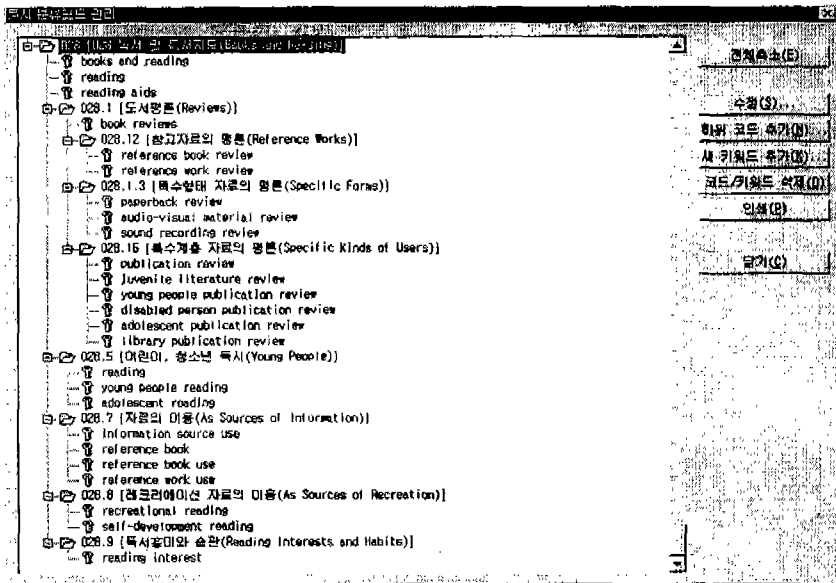
그림 6은 DDC의 '028(독서 및 독서지도)'에 대한 작업 클라이언트(그림 4)의 작업 예다. 여기에서 DDC번호 '028'은 대분류가 되고, '028.5' '028.7' '028.8' '028.9'는 중분류가 되며, '028.12' '028.13' '028.16'은 소분류가 된다.

모든 중분류가 소분류를 가지는 것은 아니다.

(b) 자동 문서분류사전의 작성 및 관리

자연언어를 처리하기 위해서는 여러 가지 방법이 사용되고 있으며 특히 사전 형태의 시스템 이용이 주류를 이루고 있다. 이러한 사전은 디지털화 되어서 전자사전을 이루게 된다. 본 연구에서도 다양한 사전을 이용하며, 그 기능과 내용을 요약하면 다음과 같다.

- ① 불용어사전: 색인어로서 가치가 없는 낱말을 수록한 사전이다. 한국어의 접속사, 영어의 접속사, 관사, 조동사 등이 수록된다.
- ② 끝말사전: 자주 사용되는 조사나 어미 등을 제거하기 위한 사전이다. 낱말의 원형을 나타내기 위하여 사용된다.
- ③ 사용자 정의사전: 사용자의 정의에 의하여 색인어 결정이 가능하게 되며, 색인의 누락을 방



<그림 6> DDC 번호 '028'에 대한 분류코드와 자동 문서분류사전 설계 예

〈표 3〉 문헌정보학분야 인터넷학술정보자원에 대한 분류코드의 설계

대분류(주제범주)	중분류(1차 하위주제)	소분류(2차 하위주제)
문헌정보학(020)	역사, 지리, 인물취급(020.9)	사서, 정보관리자(020.92)
도서관, 기록보존소, 정보센터의 관계(021)	지역과의 관계(021.2)	교육역할(021.24)
		문화역할(021.26)
		정보역할(021.28)
	도서관과 교육기관과의 관계(021.3)	없음
	협조와 네트워크(021.6)	도서관 상호간의 협력(021.64)
		도서관 네트워크, 시스템, 연합(021.65)
도서관, 기록보존소, 정보센터의 진흥(021.7)	없음	
정부와의 관계(021.8)	도서관위원회(021.82)	
	재정지원(021.83)	
물리적 시설물 관리(022)	위치, 대지(022.1)	없음
	도서관 건물(022.3)	없음
	서고와 서가(022.4)	없음
	조명(022.7)	없음
	난방, 환기, 공기조절(022.8)	없음
	비품, 설비, 장치(022.9)	없음
도서관 직원(023)	전문가 지위(023.2)	없음
	기술적 지위(023.3)	없음
	직업안내(023.7)	없음
	현직교육(023.8)	없음
	인사관리의 요소(023.9)	없음
도서관, 기록보존소, 정보센터의 운영(025)	정보의 저장과 검색(025.0)	기술과정(025.02)
		저장과 검색(025.04)
		특정분야 정보의 저장과 검색(025.06)
	도서관행정(025.1)	도서관 재정(025.11)
		복사서비스(025.12)
		특수자료(025.17)
		관종별 도서관(025.19)

( )안은 DDC 번호

〈표 3〉 문헌정보학분야 인터넷학술정보자원에 대한 분류코드의 설계(계속)

대분류(주제범주)	중분류(1차 하위주제)	소분류(2차 하위주제)
도서관, 기록보존소, 정보센터의 운영(025) -계속	수서와 장서개발(025.2)	장서개발(025.21)
		구입을 통한 수서(025.23)
		교환, 기증, 납본을 통한 수서(025.26)
		특수형태(025.28)
		지역별자료(025.29)
	서지분석과 제어(025.3)	목록(025.31)
		기술목록(025.32)
		특수자료의 목록, 분류, 색인(025.34)
		분담목록, 분류, 색인(025.35)
	주제분석과 제어(025.4)	재목록, 재분류, 재색인(025.39)
		분류와 서가목록(025.42)
		분류체계(025.43)
		주제목록(025.47)
		주제색인(025.48)
	이용자봉사(025.5)	통제어휘(025.49)
		참고봉사(025.52)
		이용자안내(025.56)
	대출봉사(025.6)	도서관 이용자연구(025.58)
		상호대차(025.62)
	물리적 처리(025.7)	없음
	장서의 유지, 보존(025.8)	장서의 배열(025.81)
장서보안, 장서점검(025.82)		
장서의 보존(025.84)		
특수도서관(026)	없음	없음
일반도서관(027)	사립, 가족도서관(027.1)	없음
	회원제도서관(027.2)	없음
	대여도서관(027.3)	없음
	공공도서관(027.4)	없음

( )안은 DDC 번호



〈표 3〉 문헌정보학분야 인터넷학술정보자원에 대한 분류코드의 설계(계속)

대분류(주제별주)	중분류(1차 하위주제)	소분류(2차 하위주제)	
일반도서관(027)-계속	국립도서관(027.5)	없음	
	특수계층을 위한 도서관 (027.6)	연령계층별 도서관(027.62)	
		소수민족 도서관(027.63)	
		특수 정부도서관(027.65)	
		사회사업체 도서관(027.66)	
		종교도서관(027.67)	
		비영리 도서관(027.68)	
		기업체 도서관(027.69)	
	대학도서관(027.7)	없음	
	학교도서관(027.8)	단계별도서관(027.82)	
종교별도서관(027.83)			
독서 및 독서지도 (028)	독서평론(028.1)	참고자료의 평론(028.12)	
		특수형태자료의 평론(028.13)	
		특수계층자료의 평론(028.16)	
	어린이, 청소년 독서(028.5)	없음	
	일반 자료의 이용(028.7)	없음	
	레크리에이션 자료의 이용(028.8)	없음	
독서흥미와 습관(028.9)	없음		

( )안은 DDC 번호

지하기 위한 것이다.

④ 자동 문서분류사전: 문서를 자동으로 분류하기 위하여 마련된 사전이다.

본 연구에서 자동 문서분류사전은 다음과 같은 과정으로 작성되었다. 우선 DDC 21판의 상관색인 편을 이용하여 대분류 8개, 중분류 39개, 소분류 49개에 해당되는 색인어를 모두 뽑아서 정리하고, DDC 21판과 함께 발행된 CD-ROM형태

의 Dewey For Windows(DFW)를 이용하여 이를 보충하였다. DDC의 책자형태를 이용한 이유는目이하의 색인어를 발견하는데 유용하고, CD-ROM형태를 여기에 병행한 이유는 책자형태에서 간과된 색인어를 찾는 데 도움이 되기 때문이다. DFW는 특정 주제의 탐색에 다양한 접근 방법을 제공하며, 그 중에서도 색인의 창(index window)은 특정주제와 관련된 색인을 조사하는데 유용하게 이용된다. DDC의 상관색인을 이용

[표 4] DDC 상관색인의 자동 문서분류사전 적용 예

하이픈(-)	단어속에 하이픈(-)이 있는 경우 하이픈으로 연결된 두 단어를 하나로 취급할 것인지, 두 단어를 분리하여 취급할 것인지를 결정해야 했으며, 본 연구에서는 하이픈을 불용어인 'and' 로 인식하게 하여 분리 취급하도록 하였다.
영어단어의 단·복수	DDC 상관색인의 경우, 영어의 명사 단어가 복수 취급되고 있었으므로 복수단어 뿐만 아니라 단수 단어도 함께 처리할 수 있도록 하였다.
영문 대·소문자	DDC 상관색인에 나오는 영문자의 대·소문자는 구분 없이 사용하였다.
접속사, 전치사	DDC 상관색인에 나오는 접속사나 전치사는 불용어로 취급하여, 이들을 생략하였다.
괄호( )	DDC 상관색인의 괄호안의 내용은 무시하였다.
약자	DDC 상관색인에서 나오는 영어약자의 경우, 약자와 전체이름 모두를 처리 하도록 하였다. (예, ISBN의 경우 ISBN과 International Standard Book Number를 다 처리)
중복	뜻은 같으나 여러 가지 형식으로 표현된 말은 그 중 하나만을 선택하여 사용하였다. (예, information retrieval과 retrieval of information의 경우 information retrieval 만 선택)
불필요한 단어의 삭제	DDC 상관색인에 나오는 불필요한 단어는 삭제하였다. 예를 들어 문헌정보학 분야의 자동 문서분류사전을 설계함에 있어 library science, information science, libraries와 같은 단어들은 꼭 필요한 경우를 제외하고는 삭제하였다.

하여 문헌정보학분야의 자동 문서분류사전을 작성함에 있어서 표 4의 규칙을 따랐다.

#### 4.3.4 URL 정보의 수동분류

로봇 에이전트에 의한 자동분류는 일반적으로 데이터의 신뢰성이 떨어진다고 생각되므로 관리자에 의한 수동분류를 자동분류와 함께 채택한다. 즉, 작업 클라이언트의 주 모듈(그림 4)을 이용하여 자동으로 분류 확정된 URL 정보라 하더라도 관련성이 없다고 판단되는 URL 정보는 수동으로 삭제할 수 있도록 하였으며, 자동으로 분류 확정되지 않은 자료라도 Dragdrop 방식 혹은

입력방식으로 관련 분류코드에 수동으로 이동시킬 수 있도록 설계하였다. 예를 들어, 작업 클라이언트(그림 4) 왼쪽 화면의 분류 중에 있는 각 URL 정보는 DragDrop 방식으로 오른쪽화면의 해당 분류코드로 이동될 수 있다.

#### 4.4 검색 클라이언트의 설계

이용자가 검색 클라이언트(WWW 브라우저)에서 DDC 21판의 소분류(2차 하위주제) 028.12 (참고자료의 평론, Reviews of Reference Works)와 그 이하에 해당되는 인터넷정보자원을

찾기까지의 과정을 예시하면 다음과 같다.

① 표 5는 초기화면으로 문헌정보학분야의 8개 대분류(주제범주)를 나타내고 있다. DDC 분류번호 '028'을 선택한다.

〈표 5〉 문헌정보학분야의 대분류

- ▶ 020 문헌정보학 (Library and Information Sciences)
- ▶ 021 도서관, 기록보존소, 정보센터의 관계 (Relationships)
- ▶ 022 물리적 시설물의 관리 (Physical Facilities)
- ▶ 023 인사 관리 (Personnel Administration)
- ▶ 025 도서관, 기록보존소, 정보센터의 운영 (Operations)
- ▶ 026 특수 도서관, 기록보존소, 정보센터 (Special Libraries)
- ▶ 027 일반 도서관, 기록보존소, 정보센터 (General Libraries)
- ▶ 028 독서 및 독서지도 (Books and Reading)

② DDC 분류번호 '028(독서 및 독서지도)'의 선택결과 표 6이 나타나며, 5개의 중분류(1차 하위주제)가 표시된다. DDC 분류번호 '028.1'을 선택한다.

〈표 6〉 문헌정보학분야 '028'의 중분류

- ▶ 028 독서 및 독서지도 (Books and Reading)
  - ▶ 028.1 도서평론 (Reviews)
  - ▶ 028.5 어린이, 청소년 독서 (Young People)
  - ▶ 028.7 일반 자료의 이용 (As Sources of Information)
  - ▶ 028.8 레크리에이션 자료의 이용 (As Sources of Recreation)
  - ▶ 028.9 독서흥미와 습관 (Reading Interests and Habits)

③ DDC 분류번호 '028.1(도서평론)'의 선택 결과, 표 7이 나타나며, 3개의 소분류(2차 하위주제)가 표시되고, DDC 분류번호 '028.12'를 선택하면 '참고자료의 평론'에 해당되는 관련 인터넷정보자원을 얻을 수 있다.

〈표 7〉 문헌정보학분야 '028'의 소분류

- ▶ 028 독서 및 독서지도 (Books and Reading)
  - ▶ 028.1 도서평론 (Reviews)
    - ▶ 028.12 참고자료의 평론 (Reference Works)
    - ▶ 028.13 특수형태 자료의 평론 (Specific Forms)
    - ▶ 028.16 특수계층 자료의 평론 (Specific Kinds of Users)

## 5. 결론

최근 WWW가 급격히 발전하고, 인터넷상에 산재한 정보의 양이 급증하면서, 이용자들은 원하는 정보를 찾는데 어려움을 겪고 있다. 이 문제의 해결책으로 제시된 것이 검색엔진이다. 여러 가지 검색엔진 중 디렉토리(주제별) 검색엔진은 자료들을 주제별로 분류해 놓고, 이용자가 단계별로 최종 정보에 도달할 수 있도록 고안된 검색엔진이다. 디렉토리 검색엔진은 수집된 자료를 이용자에게 효과적으로 제시하게 되며, 이 검색엔진의 기준은 이용자가 쉽게 자료에 접근할 수 있게 하는 것이다.

본 연구는 도서관에서 오랫동안 이용해온 분류체계인 DDC를 이용하여 문헌정보학분야의 인터넷정보자원을 수집하고 검색할 디렉토리 검색엔진을 대분류, 중분류, 소분류의 3단계로 설계한

것이다. 본 연구를 위해서는 WWW공간을 순환적으로 따라가며 정보를 모아오는 로봇 에이전트 프로그램이 필요하였으며, 이를 위하여 로봇 에이전트의 동작원리를 관리자 측면에서 살펴보았다. 다음으로 로봇 에이전트가 모아온 인터넷 정보자원을 관리자가 쉽게 관리하기 위하여 로봇 에이전트 모니터링 도구(작업 클라이언트)를 설계하였다. 여기서는 ① 로봇 에이전트의 동작요청, ② 초기 URL의 지정 및 관리, ③ 분류코드의 설계, ④ 자동 문서분류사전의 작성 및 관리, ⑤ URL 정보의 수동분류기능이 논의되었다. 특히, ③의 분류코드의 설계에서는 DDC의 문헌 분류 원칙인 체계적 순서를 적용하였고, ④의 자동 문서분류사전의 작성 및 관리에서는 DDC의 상관색인을 이용하였다.

우수한 검색엔진은 디렉토리 검색엔진 외에 키워드 검색엔진을 결합한 혼합형 검색방식을 취하는 것이 보통이다. 즉, 상위개념에서 하위개념으로 디렉토리를 이동하여 검색범위를 좁혀 가면서 검색하는 주제별 디렉토리 검색방법에 검색어를 직접 입력하여 검색하는 키워드 검색방법을 병용하는 것이다. 제한된 시간과 작업여건으로 본 연구에서는 주제별 디렉토리 검색엔진의 설계만으로 만족해야 했다. 본 연구의 디렉토리 검색엔진 설계에 키워드 검색엔진의 결합은 이 분야 인터넷정보자원의 검색능력을 더욱 향상시킬 것이다. 또한, DDC를 이용한 문헌정보학 분야 인터넷정보자원 검색엔진의 원리는 다른 분야 인터넷정보자원 검색엔진의 설계에도 적용될 수 있을 것이다.

## 참 고 문 헌

- 김성훈(1996). "까치네 로봇의 원리와 서비스 형태," 「웹코리아 웹 워크샵 자료모음」, 제 4회 WWW 워크샵 자료, WWW-KR, (<http://intranet.www-kr.org/workshop/ws4/C33/kachi>).
- 박정훈(1995). "Robot Agents and Search Engine," 「웹코리아 웹 워크샵 자료모음」, 제 2회 WWW 워크샵 자료, WWW-KR, (<http://sharon.comeng.chungnam.ac.kr/~dolphin/ws2/content/TS3-4>).
- 이영식, 황호정(1996). "인터넷 한글정보검색서비스-심마니," 「웹코리아 웹 워크샵 자료모음」, 제 4회 WWW 워크샵 자료, WWW-KR, (<http://intranet.www-kr.org/workshop/ws4/C33/simmany>).
- 정영미(1995). "인터넷 학술정보자원의 분석 및 활용에 관한 연구," 「정보관리학회지」, 12(1) : 19-43.
- 최중민(1997). "에이전트의 개요와 연구방향," 「정보과학회지」, 15(3) : 7-16.
- Koch, Traugott(1997). "The Role of Classification Schemes in Internet Resource Description and Discovery," (<http://www.ub2.lu.se/desire/radar/reports/D3.2.3>).
- Koster, Martijn(1993). "Guidelines for Robot Writers," (<http://info.webcrawler.com/mak/projects/robots/guidelines.html>).
- Koster, Martijn(1995). "Robots in the Web:

- threat or treat?" (<http://info.webcrawler.com/mak/projects/robots/threat-or-treat.html>).
- Koster, Martijn(1996). "Evaluation of the Robots Exclusion," (<http://info.webcrawler.com/mak/projects/robots/eval.html>).
- Markey, K., and A. N. Demeyer(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 Online Computer Library Center, Inc., Office of Research.
- McKiernan, Gerry(1997). "Beyond Bookmarks: Schemes for Organizing the Web," (<http://www.public.iastate.edu/~CYBERSTACKS/CTW.htm>).
- Mitchell, Joan S.(1996). "The Dewey Decimal Classification at 120: Edition 21 and beyond." 「Knowledge Organization and Change: Proceedings of the 4th International ISKO Conference. 15-18 July 1996」, Washington, D.C., Rebecca Green, ed.
- Network Wizards(1998). "Internet Domain Survey, January 1998," (<http://www.nw.com/zone/WWW/report.html>).
- Svenonius, Elaine(1983). "Use of Classification in online retrieval," 「Library Resources and Technical Services」, 27(1): 76-80.
- Vizine-Goetz, Diane(1996). "Using Library Classification Schemes for Internet Resources," 「OCLC Internet Cataloging Project Colloquium Position Paper」, (<http://www.oclc.org/oclc/man/colloq/vg.htm>)