

정보 수집 에이전트를 사용한 어린이 교육 정보 검색 시스템의 설계 및 구현

Design and Implementation of an Education Information Search System for Children Using Information Gathering Agents

전 진 옥*
Jin-Wook Jun

배 인 한**
Ihn-Han Bae

요 약

특정 정보를 필요로 하는 사용자는 관련 문서를 찾기 전에 다수의 검색 엔진들을 질의할 필요가 자주 있을 것이다. 검색 엔진들의 향해 문제를 해결하기 위하여 에이전트가 사용되고 있다. 일반적으로, 에이전트는 적절할 때 또는 다른 프로그램의 요청에 따라 자동적으로 특정한 태스크를 수행할 수 있는 프로그램이다. 이 논문에서, 우리는 EduInfo4k라 부르는 정보 수집 에이전트를 사용한 어린이 교육 정보 검색 시스템을 설계하고 구현한다. 정보 수집 에이전트는 다수의 어린이 포털 웹 사이트들: 야후코리아 꾸러기, 네이버 주니어 그리고 한미르 개구쟁이를 주기적으로 방문하고, 어린이 교육 또는 학습 정보를 수집하고, 수집된 정보를 데이터베이스에 저장한다. 따라서 일반 사용자들은 일관된 사용자 인터페이스를 통하여 데이터베이스로부터 어린이 교육 정보를 편리하게 검색할 수 있다. 그 결과, 우리는 EduInfo4k가 다수의 검색엔진에서 질의 없이 통합 검색을 제공함을 알 수 있었다.

Abstract

A user with a specific information need will often need to query several search engines before finding relevant documents. To address the problem of navigating the search engines, agents are used. In general, an agent is a program that can perform a particular task automatically, when appropriate or upon request by another program. In this paper, we design and implement the education information search system for children using information gathering agent that is called EduInfo4k. The information gathering agent periodically visits several portal web sites for children: Ggureogi of Yahoo Korea, Junior of Naver and Gaegujaengi of Hanmir, collects the education or learning information for children, stores the collected information into database. Then, causal users can search the education information for children from database through a uniform user interface, conveniently. As the result, we know that *EduInfo4k* provides integrated search without query in several search engines.

1. 서 론

우리나라 인터넷 사용자수가 1994년 이후 인터넷 열풍이 불면서 매년 두 배가 넘는 폭발적인 성장세를 지속해 왔다. 한국인터넷 정보센터에서 실시된 2001년도 인터넷 사용자수 설문조사에 의하면, 인터넷 사용자수가 2000만 이상인 것으로 집계되었고, 전체 인구의 1/3 정도가 인터넷을 사

용하는 것으로 나타나 인터넷 시대가 본격적으로 열리고 있음을 보여주고 있다. 이러한 추세를 반영하여 현재 인터넷에 존재하는 웹 서버의 수는 8억 개이고 800억 개 이상의 웹 페이지가 있다 [1]. 이처럼 인터넷에서 제공되는 정보의 양은 엄청나고 다양한 정보를 가진 웹사이트들이 지속적으로 증가하고 있다.

월드 와이드 웹을 이용한 인터넷 정보검색은 이미 보편화되어 있지만 네트워크 상에 흩어져 있는 정보의 방대함으로 인해 아직 여러 문제점을 내포하고 있다. 첫째, 웹에 존재하는 정보 사이트의 수는 매우 많으며 나날이 증가되고 있다.

* 대구가톨릭대학교 교육대학원 전자계산전공(석사)
jjw50@hanmail.net

** 대구가톨릭대학교 컴퓨터정보통신공학부 교수
ihbae@catacgu.ac.kr

이런 상황에서 찾고자 하는 정보가 정확히 어느 사이트에 있는지를 알기란 매우 어렵다. 둘째, 원하는 정보를 찾기 위해 어떤 단어를 입력으로 주어야 하는지 판단하기 어렵다. 예를 들어 ‘자동차’에 대한 정보를 찾기 위해 ‘automobile’이라고 검색어를 입력하게 되면 같은 의미인 ‘car’, ‘vehicle’, ‘truck’등을 포함한 문서는 찾기 어렵다. 셋째, 입력된 검색어를 포함한 문서가 너무 많을 경우 사용자에게 어떠한 형태로 보여주어야 하는지 어려운 판단이 될 수 있다. 넷째, 사용자의 수준에 맞는 내용을 찾기가 어렵다. 예를 들어, 초등학교생이 ‘physics’에 대하여 검색하였을 경우 물리학에 대한 박사학위 논문을 가져다준다면 소용이 없을 것이다. 이러한 여러 가지 문제점을 해결해주는 것이 바로 에이전트가 필요한 이유가 될 것이다[3].

본 논문에서는 이러한 문제점을 인식하여, 사용자의 수준에 맞는 콘텐츠를 제공하고 사용자가 여러 사이트를 방문해야 하는 번거로움을 줄이기 위하여 통합 전문 검색 시스템인 정보 수집 에이전트를 사용한 온라인 어린이 교육 정보 검색 시스템을 설계하고 구현한다. 기존의 다수의 검색 사이트들은 어린이들을 위한 포털 사이트를 따로 개설하여 서비스하고 있다. 이러한 여러 개의 어린이 포털 사이트들을 방문하여 원하는 교육 정보를 검색하는 것은 불편하다. 본 논문에서는 정보 수집 에이전트가 여러 개의 어린이 포털 사이트를 방문하여 온라인 어린이 교육 정보를 수집하고, 수집한 정보를 데이터베이스에 저장한다. 그리고 사용자들은 일관된 인터페이스를 통하여 원하는 어린이 교육 정보를 편리하게 검색하는 전문 검색 시스템을 설계하고 구현한다.

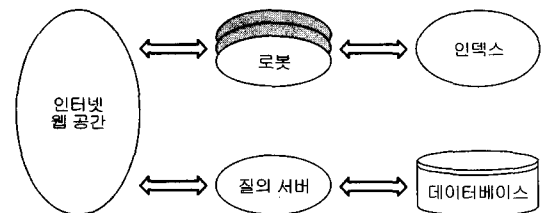
본 논문의 구성은 다음과 같다. 2장에서 관련 연구로서 인터넷과 검색엔진, 그리고 에이전트에 대하여 살펴보고, 3장에서 본 논문의 수집 에이전트를 사용한 온라인 어린이 교육 정보 검색 시스템을 설계하고, 4장에서 설계된 검색 시스템을 구현하고, 그리고 5장에서 구현된 시스템과 기존의

시스템들의 성능을 비교하고 평가한다. 그리고 마지막으로 6장에서 결론을 맺는다.

2. 관련 연구

2.1 인터넷과 검색엔진

최근 인터넷 상에서 다양한 검색 시스템이나 방대한 전문 정보 자료가 늘어나면서 이에 대한 효율적이고 효과적인 접근을 위해 여러 가지 도구들이 개발 및 사용되고 있으며, 그 중에서 웹은 현재 인터넷의 대표적인 서비스로 이용되고 있다. 이에 따라 웹을 쉽게 검색하기 위한 검색도구로서 검색엔진들이 개발되고 있고, 전 세계적으로 수많은 웹 검색엔진들이 동작하고 있다. 하지만 수많은 정보들이 널려 있기 때문에 실제로 원하는 정보를 검색하기란 쉽지 않다. 인터넷 이용자들은 그 많은 정보들 중에서 필요한 정보들을 어떻게 빠르고 간단하게 검색할 것인가 하는 문제에 직면하게 된다. 그러므로 효과적으로 원하는 정보를 수집하기 위해서는 효율적인 정보검색 방법을 숙지해야 할 필요성이 생긴다. 인터넷에서 정보를 얻는 가장 좋은 방법은 경험과 습득에 의하여 필요한 웹사이트를 북마크 해두는 방법이나 또는 신문이나 잡지 등을 통하여 웹사이트에 대한 소개를 메모해 두는 방법이다. 다른 방법으로는 유사한 주제를 다루는 사이트들에 대한 링크를 제공하는 사이트를 이용하는 방법이다. 또 다른 방법은 검색엔진 서비스를 통해서 해결책을 찾는 방법이다.



(그림 1) 검색엔진의 구조

인터넷에 흩어진 정보를 검색할 때 가장 먼저 사용하게 되는 것이 검색 엔진일 것이다. 검색 엔진은 약간의 차이점은 있겠지만 일반적으로 검색 로봇(search robot), 인덱스(index), 질의 서버(query server) 의 세 가지 구성요소를 가지고 있다(그림 1). 검색 로봇은 spider, wanderer 등으로도 불리며 주기적으로 웹 공간에 존재하는 문서를 수집하여 인덱싱 할 수 있도록 도와준다. 인덱스는 검색 로봇이 모아준 문서를 데이터베이스에 저장하는 작업을 하는데 빠르고 효율적인 검색을 위해 주로 키워드와 문서를 연결해주는 역인덱스(inverted index) 방법을 많이 사용한다. 질의 서버는 사용자의 질의 검색어를 입력으로 받아서 인덱스를 참조하여 검색 결과를 출력해준다[1]. 대부분의 검색 엔진은 로봇이 찾아온 문서의 타이틀뿐 아니라 문서 전체의 내용을 인덱싱 하는 전문 인덱싱

(full-text indexing) 방법을 택한다. 인덱싱은 특정 단어에 대해 관련 있는 문서들의 링크로 구성되는데 하나 이상의 문서가 같은 단어와 연관이 있을 경우 어느 문서가 더 관련이 있는지를 구별할 필요가 있게 된다. 이것은 사용자에게 결과를 출력해 줄 때 가장 우선순위가 높은 것부터 순서대로 보여줄 때 중요한 판단 요소가 될 수 있다. 일반적으로 검색 엔진은 동작특성에 따라 주제별 검색 엔진, 키워드 검색 엔진, 메타형 검색 엔진으로 구분된다. 표 1은 각 검색 엔진의 특징, 장·단점, 그리고 대표적인 예를 보여준다[1,2].

2.2 에이전트

에이전트는 사용자를 대신해서 사용자가 원하는 작업을 자동적으로 해결하여 주는 컴퓨터 프

(표 1) 검색 엔진의 분류

	키워드 검색엔진	제어 검색엔진
특징	로봇 에이전트를 사용하여 검색한 사이트들의 내용을 자동으로 인덱스 하여 데이터베이스를 구축하고, 사용자가 직접 질의어를 입력하여 원하는 정보를 얻는 방식이다.	하나의 사용자 입력으로부터 다수의 검색 엔진을 동시에 구동시키고, 각 검색 엔진에서 얻어진 결과를 통합하여 하나의 형태로 보여준다.
장점	정보 보유량이 많다.	- 각 검색 엔진마다 다른 사용자 입력 형태를 하나로 통합함으로써 초보자도 쉽게 이용할 수 있다. - 여러 검색 엔진을 동시에 구동시킴으로써 정보를 효율적으로 검색할 수 있다. - 여러 검색 엔진을 통하여 광범위한 검색이 가능하다.
단점	- 광범위한 단어 입력 시 너무 많은 결과를 보여 주므로 적절한 정보를 찾기 어렵다. - 연산자 및 옵션에 대한 사전 지식을 갖추어야 검색 엔진을 제대로 활용할 수 있다.	- 여러 검색 엔진을 구동시키기 위하여 더 많은 자원이 필요하다. - 각 검색 엔진에서 제공하는 인터페이스 기능을 이용하지 못한다. - 검색 요구가 많을 경우 병목현상이 발생하여 수행속도가 떨어진다. - 검색시간이 비교적 오래 걸린다.
예	Yahoo, Virtual Library, Simmani	Altavista, Lycos, Empas, Naver Metasearch, 미스다찾나

로그를 말한다. 에이전트의 개념을 정리해 보면 다음과 같이 세 가지로 요약할 수 있다.

- 에이전트는 자율적이고 목적 지향적인 프로세스이다.
- 에이전트는 어떤 환경 안에 위치하고, 그 환경의 특성을 인지하며, 환경 변화에 반응한다.
- 에이전트는 작업의 수행을 위해 다른 에이전트나 사용자와 협동한다.

에이전트의 특성을 크게 세 가지로 나타낼 수 있는데, 자율성과 협동성, 그리고 적응성이다. 자율성은 사용자나 다른 에이전트의 직접적인 지시 없이도 스스로 행동하는 성질이며 에이전트를 특징 지워주는 가장 두드러진 성질이다. 자율성에는 환경의 변화에 수동적으로 대처하는 반

응성 외에도 자신의 목적을 가지고 능동적으로 작업 수행을 추구하는 목적 지향 행동도 포함된다. 협동성은 표준 언어나 프로토콜을 이용하여 다른 에이전트와 공통된 목적을 달성하는 성질이다. 적응성은 학습할 수 있는 능력을 나타내며 환경이나 사용자에게 맞도록 적응하여 수행 효율을 높이는 성질이다. 90년대에 들어서면서 에이전트에 대한 관심과 필요성이 증대되고 있는데, 그 이유 중에는 인터넷의 폭발적인 이용 증가로 인하여 사용자들에게 원하는 정보를 편리하게 제공해 주는 기능이 요구되었기 때문이다. 즉, 인터넷의 정보 홍수 속에서 원하는 정보를 정확하게 제시기에 얻기란 쉬운 일이 아니며, 따라서 이러한 작업을 대신해주는 에이전트의 역할이 점점 커지고 있다[3].

인터넷 상에서 필요한 정보를 처리하는 기능을 가진 에이전트를 정보 에이전트라고 하는데 이

(표 2) 정보 에이전트의 분류

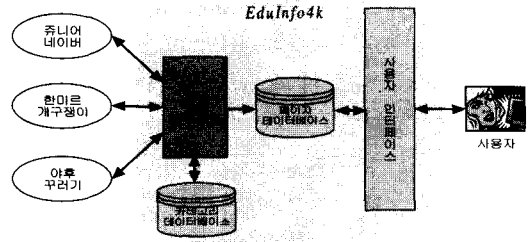
	기능	정보 검색 에이전트는 사용자가 원하는 정보를 찾아주는 역할을 수행한다.
	예	검색 엔진
	기능	정보 여과는 이 메일이나 뉴스 그룹의 정보와 같은 정보 스트림을 사용자의 프로파일과 비교하여 관심있는 정보만 걸러서 저장한 후 사용자는 걸러진 정보만을 볼 수 있다. 사용자는 정보 여과 과정을 거친 결과를 본 후 그것이 실제로 자신이 원하는 것이었는지를 알려주는 관련성 피드백을 통하여 사용자 프로파일을 재구성할 수 있다.
	예	웹 문서 필터링 에이전트, 사용 뉴스 필터링 에이전트, 유즈넷 필터링 에이전트
	기능	정보 통합 에이전트는 인터넷에서 제공되는 다수의 정보 사이트에서 사용자가 원하는 정보를 추출하여 하나의 형태로 제공하는 기능을 수행한다.
	예	비교 쇼핑 시스템
	기능	정보 수집 에이전트는 다양한 정보 소스를 사용하여 다른 에이전트에 의해 제출된 질의에 응답하는 것이다.
	예	방문객을 위하여 미팅 스케줄을 작성하는 에이전트는 그 방문객의 스케줄에 대한 부수적인 제한(예, 가용 회의실 검사, 미팅 시간 배정에서 우선순위 결정을 위한 조직도 검사)을 찾기 위하여 다수의 정보 소스들과 협의한다.
	기능	정보 추출은 한 문서에서 그 문서의 중심적 의미를 나타내는 특정 구성요소를 인식하여 추출하는 작업을 수행한다.
	예	날씨 정보를 제공하는 웹 문서에서 지역, 날짜, 온도, 습도를 추출한다.

정보 에이전트는 보통 정보검색 에이전트, 정보필터링 에이전트, 정보통합 에이전트, 정보추출 에이전트 정보 수집 에이전트 등으로 분류할 수 있다. 표 2는 각 정보 에이전트에 대한 특징을 간략히 설명한다[4].

3. EduInfo4k의 설계

인터넷이 폭발적으로 팽창함에 따라 각자가 소화해야 할 정보는 그 만큼 늘어나면서 오늘날을 사는 정보인의 부담은 점점 커지고 있다. 그 부담을 해소시킬 수 있는 정보 검색 서비스가 검색엔진이다. 하지만 사용자들은 원하는 정보를 얻기 위해 여러 사이트들을 방문하여 질의하는 번거로움과 많은 시간이 소요되는 문제점을 겪을 수 있다. 따라서 본 논문에서는 정보 수집 에이전트가 주기적으로 어린이 교육 포털 사이트들을 방문하여 어린이 학습 정보를 검색 한 후 검색된 자료를 수집하여 데이터베이스에 저장하고, 사용자는 일관된 인터페이스를 통하여 데이터베이스에 저장된 어린이 학습정보 편리하게 검색하고 획득할 수 있는 통합 전문 검색 시스템인 정보 수집 에이전트를 사용한 온라인 어린이 교육 정보 검색 시스템, EduInfo4k(Education Information for Kids)를 설계하고 구현한다.

본 논문에서 설계한 EduInfo4k는 그림 2와 같이 크게 3가지 구성요소 : 정보 수집 에이전트, 데이터베이스, 그리고 사용자 인터페이스로 구성 되어있다. 여기서 정보 수집 에이전트는 세 개의 어린이 포털 사이트: 야후 꾸러기, 네이버 주니어, 한미르 개구쟁이를 방문하여 사이트들의 카테고리를 검색 한 후 온라인 교육/학습 자료(카테고리, 사이트 주소)를 수집하여 데이터베이스에 저장한다. 그리고 사용자 인터페이스는 사용자와 Eduinfo4k간의 인터페이스 역할을 한다. 사용자 인터페이스는 일관된 질의 인터페이스를 제공하고 일관된 질의 결과를 출력시킨다.



(그림 2) EduInfo4k의 구조

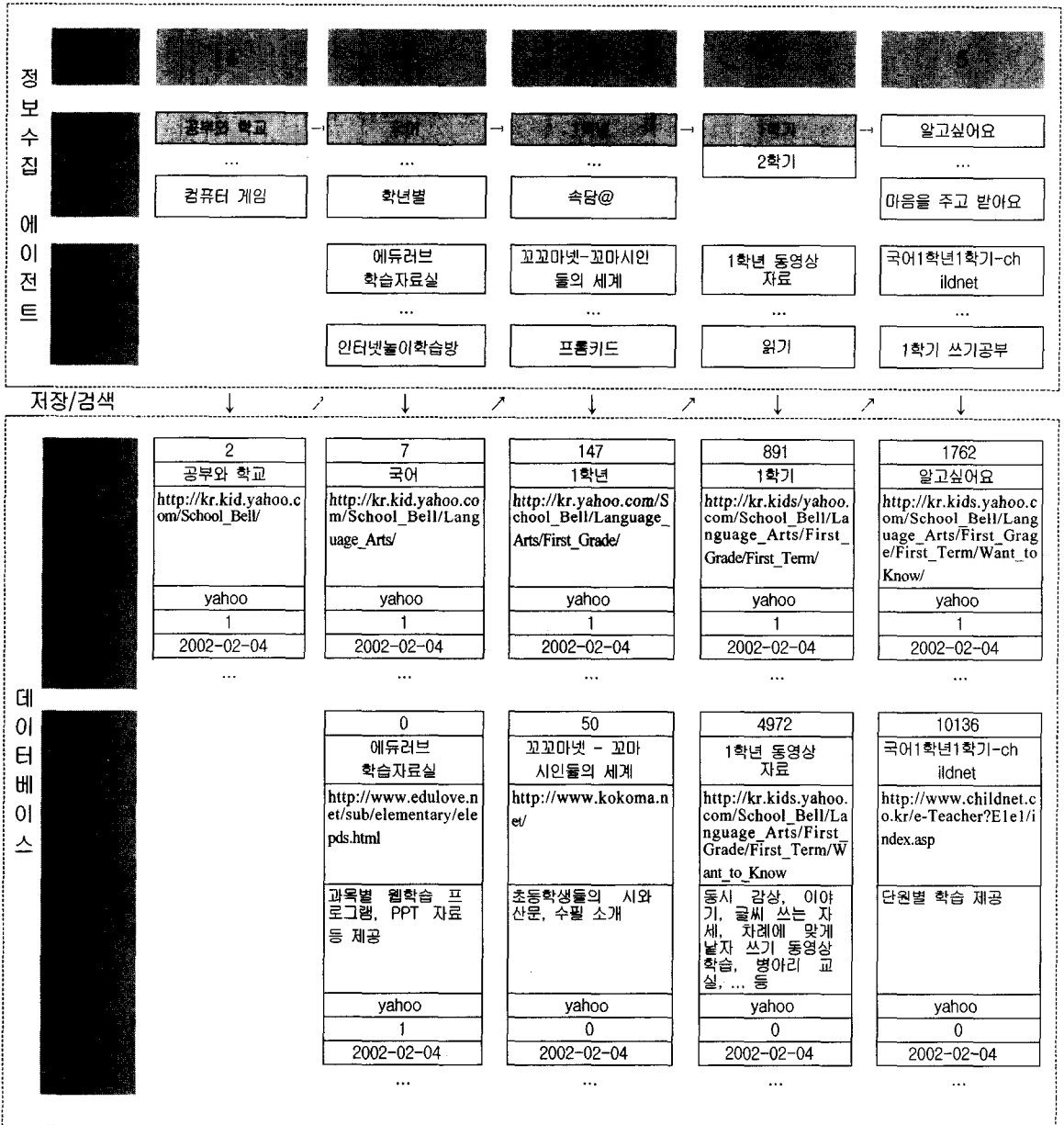
Eduinfo4k의 실행과정은 다음과 같다. 먼저 정보 수집 에이전트는 야후 꾸러기 사이트를 방문하여 카테고리(처음엔 카테고리만 존재)를 검색한 후 카테고리 정보를 카테고리 데이터베이스에 저장한다. 그리고 그 카테고리과 링크된 다음 단계의 카테고리도 이동하면 역시 카테고리과 웹 페이지가 존재하는데, 정보 수집 에이전트는 그 카테고리과 웹 페이지 정보를 카테고리 데이터베이스와 페이지 데이터베이스에 각각 저장한다. 이런 방식으로 5 단계까지의 카테고리 정보와 웹 페이지 정보를 수집한다. 이 단계가 끝나면 정보 수집 에이전트는 다음 포털 사이트인 네이버 주니어, 한미르 개구쟁이를 방문하여 위와 같은 과정을 반복한다. 그리고 사용자와 사용자 인터페이스의 질의창을 통하여 검색키를 입력하거나 검색 키워드를 클릭 하게되면 사용자 인터페이스는 검색키를 입력으로 질의어를 생성하여 데이터베이스를 검색하고, 그 결과를 일관된 형식으로 보여준다.

(표 3) 페이지 데이터베이스 테이블의 구조

항목	자료형	내용
num	int	번호
title	varchar	제목
url	varchar	주소(사이트 중복 여부, 기본키)
cont	text	주석
site	varchar	포털 사이트명
ox	varchar	카테고리 방문 여부
date	datetime	갱신날짜

Eduinfo4k에서 웹 페이지 정보를 저장하는 페이지 데이터베이스의 테이블 구조는 표 3과 같다. 여기서 num는 자료의 추가, 삭제 및 수정이 용이하게 하는 일련번호이고, title에서는 웹 페이지들의 제목을 저장하게 된다. 그리고 url은 웹 페이지들의 url(주소)를 저장하게 되는데, 이는 기존에 방문하여 저장된 url과 비교하여 이미 방문한 url이면 스킵하고 그렇지 않으면 url 정보를 저장한다. 따라서 url은 사이트 중복여부를 체크하는 기본 키로 사용된다. cont는 웹 페이지들의 주석을

지들의 url(주소)를 저장하게 되는데, 이는 기존에 방문하여 저장된 url과 비교하여 이미 방문한 url이면 스킵하고 그렇지 않으면 url 정보를 저장한다. 따라서 url은 사이트 중복여부를 체크하는 기본 키로 사용된다. cont는 웹 페이지들의 주석을



(그림 3) 정보 수집 에이전트의 실행 과정

저장하고, site는 웹 페이지의 정보를 가져오기 위해 방문한 포털 사이트 명을 저장한다. 그리고 ox는 카테고리의 방문여부를 판단하여 이미 방문한 카테고리 정보는 스킵을 하고 다음 카테고리를 검색하게 된다. 마지막으로 date는 정보를 저장한 갱신날짜를 저장한다. 그리고 카테고리 데이터베이스 테이블은 페이지 데이터베이스 테이블에서 웹 페이지 주석을 저장하는 cont 항목을 제외한 항목들로 구성된다.

그림 3은 Eduinfo4k의 정보 수집 에이전트가 야후 꾸러기 사이트를 방문하여 교육 정보를 수집하는 과정을 보여준다. 먼저 정보 수집 에이전트는 1 단계 카테고리 정보인 ‘공부와 학교’ 등을 수집하여 카테고리 데이터 베이스에 레코드 (2, 공부와 학교, http://kr.kid.yahoo.com/School_Bell/yahoo, 1, 2002-02-02)’ 등을 저장한다. 카테고리 정보 ‘학교와 공부’를 이용하여 2 단계 카테고리 정보인 ‘국어, ..., 학년별’과 웹 페이지 정보인 ‘에듀러브 학습자료실, ..., 인터넷놀이학습방’을 수집하여 카테고리 데이터베이스에 레코드 (7, 국어, http://kr.kid.yahoo.com/School_Bell/Language_Arts/yahoo, 1, 2002-02-04) 등과 페이지 데이터베이스에 레코드 (0, 에듀러브 학습자료실, <http://www.edulove.net/sub/elementary/elepds.html>, ‘과목별 웹학습 프로그램, PPT 자료 등 제공’, yahoo, 1, 2002-02-04) 등을 각각 저장한다. 이런 과정을 5 단계 카테고리까지 계속한다. 그리고 사용자가 ‘검색 키워드를 클릭하거나 입력하면, 페이지 데이터베이스에서 질의어를 수행하여 그 결과를 일관된 형식으로 사용자에게 보여준다.

4. Eduinfo4k의 구현

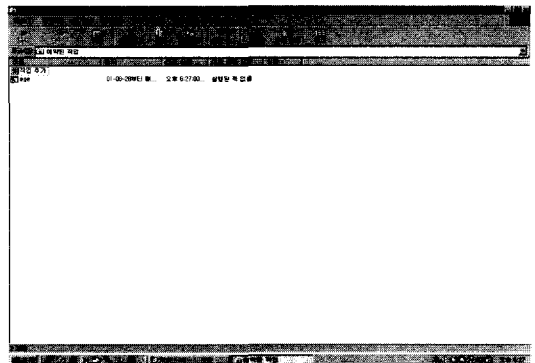
4 장에서 정보 수집 에이전트를 사용한 어린이 교육 정보 검색 시스템인 Eduinfo4k의 구현 내용과 실행 결과에 대하여 설명한다. 본 논문에서 구현한 Eduinfo4k의 개발 환경은 표 4와 같다.

(표 4) Eduinfo4k의 개발 환경

항목	사양
프로토콜	TCP/IP
운영체제	Window98
웹서버	Apache 1.3
DB 서버	My SQL 3.23.11-alpha
사용언어	PHP, Java Script
URL	http://eduinfo4k.wo.to

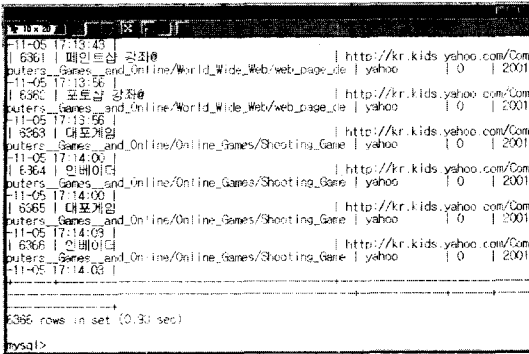
기존의 웹 문서들을 수집하기 위한 웹 로봇, 에이전트들은 C와 같은 고급언어와 복잡한 알고리즘으로 구현되었지만 Eduinfo4k는 간단한 스크립트 언어인 PHP와 자바 스크립트로 구현되었다. 그리고 아파치 서버에 MYSQL을 사용하여 소형 시스템에서도 충분히 실행되도록 하였다.

그림 4는 정보 수집 에이전트가 실행되는 과정을 보여준다. 정보 수집 에이전트를 주기적으로 실행시키기 위해 간단히 윈도우98에 내장된 작업 스케줄러를 이용하였다. 정보 수집 에이전트는 일주일에 한번씩 주기적으로 포털 사이트에 방문하여 정보를 수집하도록 하였다.



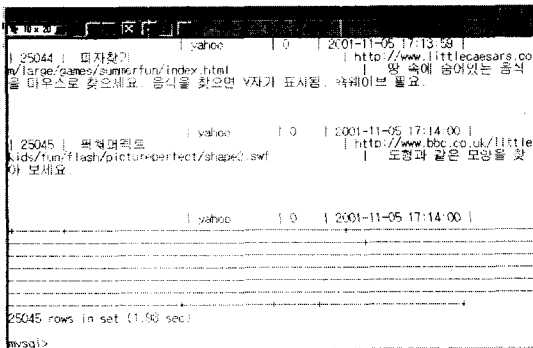
(그림 4) 정보 수집 에이전트의 실행

그림 5는 카테고리의 정보들이 저장되는 과정을 보여준다. 카테고리의 제목과 url 정보, 그리고 그 카테고리가 속한 포털 사이트 명과 갱신날짜가 저장됨을 알 수 있다.



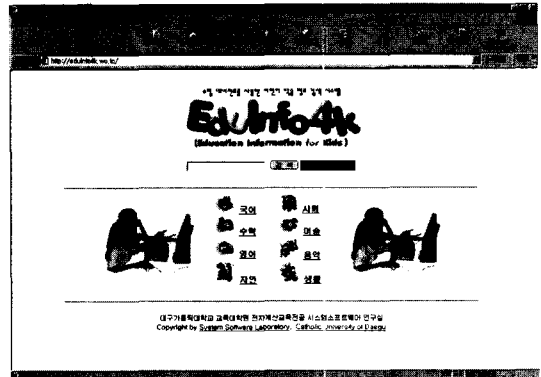
(그림 5) 카테고리 정보 저장

그림 6은 웹 페이지들의 정보가 저장되는 것을 보여준다. 웹 페이지들의 제목과 url, 그리고 웹 페이지들에 해당되는 주석, 웹 페이지들이 속한 포털 사이트 명, 마지막으로 갱신날짜가 저장됨을 볼 수 있다.



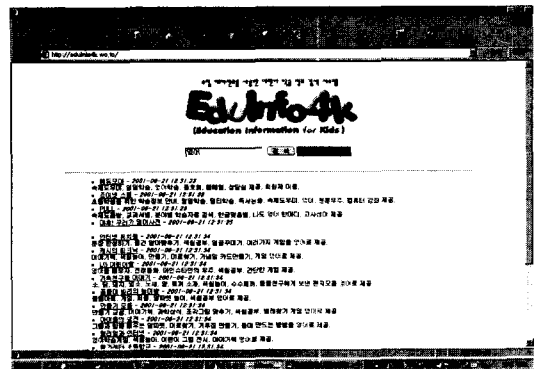
(그림 6) 웹 페이지 정보 저장

그림 7은 구현된 Eduinfo4k의 초기실행 화면을 보여준다. 초기화면에 입력창을 두어 사용자가 검색 키워드를 입력할 수 있도록 하였고, 그 옆에는 간단한 도움말을 두었다. 그리고 초등학교 교과목 명 및 자주 사용되는 질의를 초기화면에 배치하여 검색 키워드로 사용할 수 있도록 하였다. 아울러 본 논문의 Eduinfo4k가 어린이 교육 정보를 제공함으로써 초기 화면을 어린이에게 친숙한 이미지와 아이콘으로 구성하였다.



(그림 7) Eduinfo4k의 초기화면

그림 8은 초기화면에 나열된 검색 키워드 중의 하나인 '영어'를 클릭 하여 검색된 결과를 보여준다. 검색된 교육 정보는 해당 사이트로 링크되어 있으며 파란색으로 표시되고, 검색된 교육 사이트에 대한 주석에서 해당검색 키워드는 붉은 색으로 표시된다.

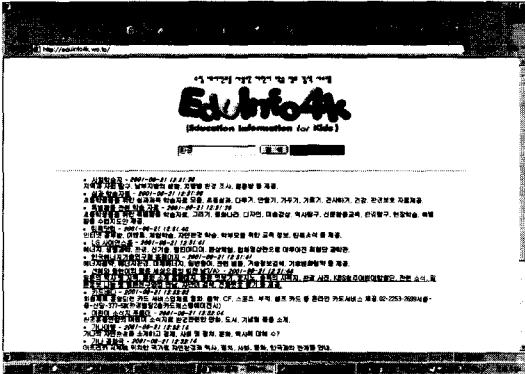


(그림 8) 검색 키워드를 통한 검색 결과

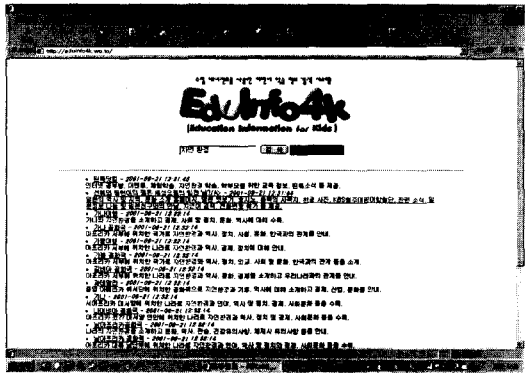
그림 9는 입력창에 검색 키워드로 '환경'을 입력한 경우 Eduinfo4k에 의해 검색된 결과를 보여준다.

Eduinfo4k에서는 부울 연산자를 사용한 AND와 OR 검색기능을 제공한다. AND 연산일 경우 하나의 키워드를 입력한 다음, 하나의 공백을 띄우고 다음 키워드를 입력하면 두 키워드 모두 존재

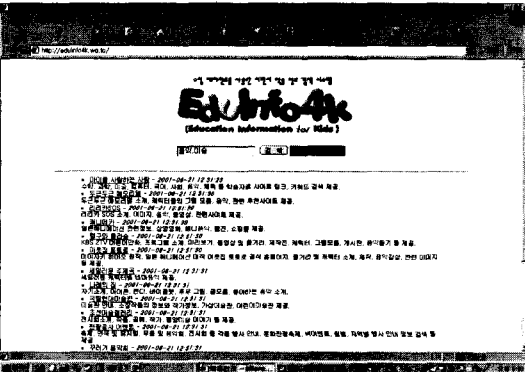
하는 웹 페이지 검색이 이루어진다. 그림 10은 검색창에 '자연 환경'을 입력하여 AND 검색된 결과를 보여준다.



(그림 9) 입력창을 통한 검색 결과



(그림 10) AND 연산 검색 결과



(그림 11) OR 연산의 검색 결과

OR 연산인 경우 키워드 사이에 콤마(,)를 입력하면 두 키워드 중에서 어느 하나의 키워드가 존재하는 웹 페이지 검색이 이루어진다. 그림 11은 검색창에 '음악, 미술'을 입력하여 OR 검색된 결과를 보여준다.

5. 평가

성공적인 검색 엔진이 갖추어야 할 요건을 간단히 정리하면 다음과 같다.

- 많은 정보를 가지고 있을 것.
- 데이터베이스의 내용 및 갱신(수정) 주기가 빠를 것.
- 검색 속도가 빠를 것. 넷째, 검색옵션의 다양성 및 편리한 사용자 환경(interface)을 제공할 것.

검색엔진의 주목적은 누가 뭐래도 정확한 검색이다. 이를 위해서는 여러 검색 옵션이 제공되어야 한다. 각 검색엔진들은 다양한 검색 조건을 이용한 복합 검색 기법을 활용하고 있다. 부울(boolean) 연산은 AND, OR, NOT과 이의 조합뿐 아니라 비슷한 단어까지 함께 검색하는 유의어 검색이 가능하고, '한글*' 이라고 입력하면 '한글'로 시작하는 모든 단어를 찾아주는 절단 검색 기능도 지원된다. 특히 불린 연산자의 한글 표현도 가능한데, 'AND'의 경우 '&' 뿐 아니라 '와', '과', '및' 등을 사용할 수 있다.

완벽한 검색을 수행했다라도 그 결과가 보기 쉽게 나타나야 원하는 정보를 파악하기 쉽다. 보통 분류 검색이 로봇 검색에 비해 나은 장점은 사람이 분류해 놓았기 때문에 간략한 설명이라도 볼 수 있다는 점이다. 또한, 검색 결과를 한 화면에 몇 개까지 나타낼 것인가, 분류 항목과 검색 결과가 함께 표시되는가 하는 점 등도 주요 체크 포인트이다.

(표 5) 어린이 교육 정보 검색 시스템들의 성능 평가

	http://kr.yahoo.com	http://www.naver.com	http://www.hanmir.com	EduInfo4k http://eduinfo4k.wo.to
	××	○	○	○
	×	○	○	×
	×	×	×	×
	and, or	and, or, not	and, or, not	and, or
	×	○	○	○
	×	×	×	○
	○	○	○	○
	○	○	○	×
	○	○	○	○

검색엔진들의 디자인상 가장 큰 문제점은 과도한 그래픽 사용이었다. 특히 상용화가 추진되면서 각 엔진마다 광고를 싣고 있는데, 이들 광고는 로드 속도를 더디게 하는 주원인이 되고 있다. 가장 유명한 검색엔진인 야후와 알타비스타의 화면이 간결하다 못해 썰렁하기까지 한 이유는 검색엔진의 생명은 빠른 속도라는 것을 잘 인식하고 있기 때문이다.

본 논문의 Eduinfo4k와 기존의 어린이 교육 정보 검색 시스템과의 기능별 성능 평가 결과는 표 5와 같다. 정보 수집 에이전트를 사용하고 어린이 교육을 위한 전문검색이 가능하고, 여러 사이트를 방문하여 자료를 수집하여 결과를 보여주는 통합 검색이 가능하고, 그리고 데이터베이스에 저장된 자료를 보여주기 때문에 검색속도 역시 빠르다. 아울러 일관된 인터페이스로 사용자에게 편리한 환경을 제공한다. 그러나 Eduinfo4k는 전문검색 시스템을 지향하기 때문에 포털기능은 제공하지 않는다.

6. 결 론

최근 인터넷의 급속한 성장과 보급으로 이메일이나 웹과 같은 정보 스트림을 통해 제공되는 정

보의 양은 엄청나며, 그 종류도 뉴스 정보, 상품 카탈로그, 영화 스케줄 등 매우 다양하다. 하지만 이처럼 웹을 통해 접근 가능한 테라바이트 수준의 정보량에 비해 실제 사용자 개개인이 필요로 하는 정보는 극히 일부분이며, 이러한 현상을 정보과다(information overload)라고 표현할 수 있다. 이 정보 과다로 인해 생산성 증대, 교육적 이득, 오락적 가치 등 인터넷을 이용하는 이점이 위협받을 수 있다. 이러한 현실 속에 인터넷의 많은 정보 중에서 필요한 정보만을 사용자에게 제공할 수 있는 에이전트의 필요성은 날이 증대되고 있다. 본 논문에서는 정보 수집 에이전트가 어린이 교육 포털 사이트들을 주기적으로 방문하여 어린이 교육 정보를 수집하고, 수집한 정보를 데이터베이스에 저장하고, 그리고 사용자들은 일관된 인터페이스를 통하여 데이터베이스에 저장된 인터넷 어린이 교육 정보를 편리하게 검색하고 획득할 수 있는 EduInfo4k를 설계하고 구현하였다. EduInfo4k는 통합 전문 검색이 가능하여 사용자가 원하는 정보를 찾기 위하여 일일이 어린이 교육 포털 사이트들을 항해해야 하는 번거로움이 해소될 뿐만 아니라 수준별 검색도 가능하였다. 향후 연구 과제로는 EduInfo4k에 수집된 정보에 대한 신뢰성 검사 기능과 사용자 프로파일에 기

초한 정보 필터링 기능을 추가하는 것과 정보 수집 에이전트를 사용한 어린이 교육 매체 가격 비교 검색 시스템을 개발하는 것이다.

참 고 문 헌

- [1] 최중민, “인터넷 정보 가공을 위한 에이전트 연구동향”, <http://cse.hanyang.ac.kr/~jmchoi/papers/info-gatherig/nfo-gathering.tml>
- [2] 홍재현, “WWW 검색엔진의 특징 및 기능 분석”, 중부대학교 논문집, 제 12권, 1999.
- [3] 장영범, “소프트웨어 에이전트 개요”, <http://user.chollian.net/~keyman21/cp/swagent.html>
- [4] 최중민, “인터넷 정보추출 에이전트”, 한국정보과학회 학회지, 제 18권 5호, pp. 48-53, 2000.
- [5] 승현석, “정보검색의 새로운 세계, 에이전트를 잡아라”, <http://www.mochanni.com/~hseung/writing/infoage/9604/article.html>
- [6] 윤효근, 이희국, 이상용, “사용자 행위에 기반을 둔 적응형 신문기사 수집 에이전트 시스템”, 한국인터넷정보학회, 춘계학술발표논문집, 제 2권 1호, pp. 208~212, 2001.
- [7] 하은용, 정명교, “MySQL과 PHP를 이용한 Internet 가격 비교 검색엔진의 설계 및 구현”, 한국정보처리학회 추계학술발표논문집 제 7권 2호, pp. 1493~1496, 2000.
- [8] Alon y. Levy, Yehoshua Sagiv, Divesh Srivastava, “Towards Efficient Information Gathering Agents”, AAAI Spring Symposium on software Agents, pp. 64~70, 1994
- [9] 이연후, “인터넷 검색엔진”, <http://www.netro/yllee/engin04.htm>
- [10] 문승주, “각 엔진의 비교 및 성능분석”, http://www.netusers.org/info_search/mun_engine_09.html
- [11] 고일석, 완전초보자를 위한 자바스크립트, pp. 639, 가남사, 2000.
- [12] Castagnetto 외 4명, PHP programming, pp. 980, 정보문화사, 2000.
- [13] 전진욱, 김현정, 배인한 “정보 수집 에이전트를 사용한 어린이 온라인 교육 정보 검색 시스템의 설계 및 구현”, 한국인터넷정보학회 추계학술발표대회 논문집, 제 2권 2호, pp. 99~102, 2001.

● 저자 소개 ●

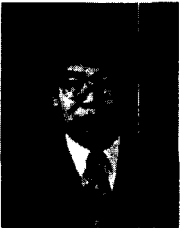


전진욱

1999년 2월 경일대학교 컴퓨터공학과(학사)

2002년 2월 대구가톨릭대학교 교육대학원 전자계산전공(석사)

E-mail : jjw50@hanmail.net



배인한

1984년 2월 경남대학교 전자계산학과(학사)

1986년 2월 중앙대학교 대학원 전자계산학과(석사)

1990년 8월 중앙대학교 대학원 전자계산학과(박사)

1996년~1997년 Computer and Information Science, The Ohio State University(Post-doctoral)

2002년~현재 : Computer Science, Old Dominion University(Visiting scholar)

1989년~현재 : 대구가톨릭대학교 컴퓨터정보통신공학부 교수

E-mail : ihbae@cataegu.ac.kr