

다중 웹 게시판에서의 정보수집 및 관리 시스템 설계 및 구현

The Design of Information Collection and Management System in Multiple Web Board

김광식, 이승관*, 정태충**
경희대학교, 가톨릭대학교*, 경희대학교**

Kim Kwang-Sik, Lee Seung-Gwan*,
Chung Tae-Choong**
KyungHee Univ., Catholic Univ.*,
KyungHee Univ.**

요약

인터넷의 보급과 발전으로 인해 공유하고자 하는 정보를 수시로 웹 브라우저상의 간단한 조작만으로 갱신 및 공유할 수 있고, 정보를 누적 관리할 수 있는 웹 게시판(web board)은 단연 웹상의 정보공유의 중요한 매체로 자리 잡고 있다. 하지만 게시판 정보들은 사용자에게 통보 없이 수시로 변하고 있어서 사용자가 개별적으로 게시판을 방문하여 일일이 확인하고 검색하여 필요한 정보를 얻는다는 것은 매우 어려운 일이다. 따라서 효율적이고 체계적인 게시판 정보관리와 검색을 위한 기술들이 개발되고 향상되고 있는 현실에 맞추어 본 논문에서는 기존의 다중게시판에서 정보 관리와 검색이 불편하던 점을 개선하여 원하는 정보를 빠르고 편리하게 모니터링 및 검색, 관리할 수 있는 다중게시판 사용자 인터페이스를 설계 및 구현을 한다.

I. 서론

최근 인터넷의 급속한 성장과 보급으로 이메일이나 웹과 같은 정보 스트림을 통해 제공되는 정보의 양은 엄청나며, 그 종류도 뉴스 정보, 구인구직정보, 상품 카탈로그 등 매우 다양하다. 하지만이처럼 웹을 통해 접근 가능한 테라바이트 수준의 정보량에 비해 실제 사용자 개개인이 필요로 하는 정보는 극히 일부뿐이며, 이러한 현상을 정보과다(information overload)라고 표현할 수 있다. 이 정보 과다로 인해 생산성 증대, 교육적 이득, 오락적 가치 등 인터넷을 이용하는 이점이 위협을 받을 수 있다. 이러한 현실 속에 인터넷의 많은 정보 중에서 필요한 정보만을 사용자에게 제공할 수 있는 에이전트의 필요성은 날이 증대되고 있다[1]. 따라서 본 논문에서는 다중 웹 게시판의 주소와 약간의 정보를 미리 등록해 주는 것만으로 여러 게시판들 속에서 자기가 원하는 최신정보들을 실

시간으로 수집하고 게시판별 검색 또는 통합검색을 할 수 있으며 간단한 클릭만으로 수집한 정보를 원하는 형태로 정렬하여 열람 및 관리할 수 있고 환경설정 등을 통하여 등록한 키워드에 해당하는 정보는 매일로 받아볼 수 있게 하는 "다중 웹 게시판 정보수집 및 관리시스템"을 설계 하고자 한다. 이것은 매일 수많은 게시판을 방문하여 정보갱신을 확인하고 필요한 정보를 수집하는데 낭비되는 시간을 절감시킬 수 있으며 효과적인 인력 운영으로 다중 게시판의 정보수집에 필요한 유지비용 절감도 예상할 수 있다.

II. 관련연구

2.1 에이전트의 연구

이 논문에서 논의되는 에이전트(agent)는 기본적

으로 소프트웨어(software) 에이전트[2]로써 큰 범주로 보면 일종의 프로그램(program)이라 볼 수 있다. 물론, 로봇과 같은 하드웨어 에이전트도 에이전트의 범주에 포함시키기도 한다. 그렇지만 이때 대두되는 질문은 과연 소프트웨어 에이전트라는 것이 기존의 일반 프로그램과 무엇이 다르냐 하는 것이다. 바꾸어 표현하면 기존의 소프트웨어(예를 들어, 워드 프로세서, 스프레드 시트, 게임 등)들도 모두 에이전트라고 할 수도 있지 않느냐 하는 것들이다. 이를 위해서 단순한 프로그램이 아닌 소프트웨어 에이전트이기 위해서는 어떤 기능을 수행하고 어떤 역할을 담당해야 하는가에 대한 답변이 요구되는데 이에 대한 답변이 바로 에이전트의 정의가 될 수 있다[3].

2.2 웹 게시판 정보수집 방법

1) 일반적인 웹 게시판 정보수집 방법

일반적인 웹 게시판 정보수집 방법으로는 사용자가 목표 웹 게시판들을 일일이 방문하여 정보를 수집하는 방법이다. 이때 다중 게시판에서 갱신된 게시물은 사용자가 미리 인지하지 못하므로 모든 게시판을 방문하여야만 갱신된 정보를 확인 및 수집할 수 있기에 인력과 시간 자원의 낭비가 심하다.

2) 목표 웹 게시판들의 데이터베이스에 접근하여 정보를 수집하는 방법

에이전트가 직접 목표 웹 게시판 데이터베이스에 접속하여 최근 갱신 정보를 확인하여 로컬데이터베이스에 저장했다가 다시 사용자 인터페이스를 통하여 사용자한테 모니터링을 도모하는 방법이다.

3) 웹 게시판의 URL에 접속하여 HTML코드를 파싱(Parsing)하여 정보를 수집하는 방법

게시판 URL에 접속하여 HTML코드를 분석하여 최근갱신 정보만 추출함으로써 목표 게시판의 데이터베이스 구조나 데이터베이스유형에 구애 받지 않고 동일 네트워크상의 게시판이나 외부의 게시판이

나 상관 없이 게시물을 수집하여 로컬 데이터베이스에 저장하여 검색 및 관리하는 방법이다.

Ⅲ. 다중 웹 게시판 정보수집 및 관리시스템

3.1 시스템 구현시 고려사항

자동으로 다중 웹 게시판 정보를 수집하여 다중 사용자에게 검색 및 관리 그리고 특정 서비스를 지원하기 위하여 아래의 사항이 고려되어야 한다.

1) 웹 기반 인터페이스

다중 웹 게시판 정보수집 및 관리시스템의 목적은 편리하고 신속하게 게시판들의 정보를 전달하는데 있다. 따라서 간단한 조작만으로도 충분히 시스템의 모든 기능을 사용할 수 있도록 편리한 구성이어야 한다. 가급적 클라이언트에 설치해야하는 소켓 프로그램이나 애플릿 등은 제외하고 서버 측에서 모든 기능을 할 수 있도록 제작하여야 한다[4]. 따라서 논문에서는 웹을 통하여 사용 가능한 시스템을 설계하였다.

2) HTML페이지 파싱(Parsing)을 통한 정보수집

목표 웹 게시판들의 데이터베이스에 접근하여 정보 수집 하는 방법에는 아래와 같은 두 가지 문제점이 존재한다.

문제점1 : 목표 웹 게시판들의 데이터베이스가 동일하지 않을 경우에 에이전트는 게시판들의 다양한 유형의 데이터베이스에 접속을 해야 하고 또한 웹 게시판 데이터베이스마다 테이블 구조가 일정하지 않기에 정확히 필요한 정보를 추출하기엔 무리가 따른다.

문제점2 : 동일네트워크일 경우에는 데이터베이스 접근을 허용할 수도 있지만, 목표 웹 게시판의 데이터베이스가 원격 네트워크일 경우에는 접근을 위하

여 데이터베이스의 포트를 외부에서 접근이 가능하도록 오픈하여야 한다. 목표 웹 게시판 관리자는 보안상 데이터베이스 포트를 오픈하지 않을 것이며 접속 정보를 제공하지 않을 것이다.

따라서, 본 논문에서는 데이터베이스로의 접근이 아닌 직접 게시판들의 웹 URL에 접속하여 HTML 페이지를 파싱(Parsing)하여 필요한 정보를 수집하려고 한다.

3) 정보의 누적관리

웹상의 정보는 수시로 갱신되며 또한 수시로 소실되고 있다. 때문에 수집한 정보를 로컬데이터 베이스에 저장하여 자신만의 정보 데이터베이스를 구축하여 소장할 필요가 있다.

4) 정보관리의 편리성

정보를 수집하여 로컬데이터 베이스에 저장한다고 하여 쉽게 자기가 원하는 정보를 찾아낼 수 있는 것은 아니다. 따라서 게시판별, 제목별, 작성자별로 검색이 가능해야 하며 원하는 대로 정렬하여 열람할 수 있어야 한다.

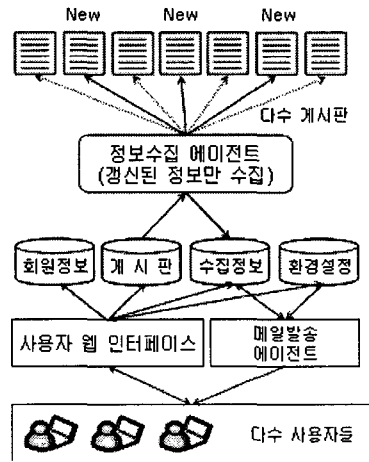
간혹 관리시스템에 접속이 불가능 할 경우 메일 혹은 휴대폰으로의 실시간 정보 입수도 필요하다.

5) 구현비용 저렴

만일 비싼 비용을 들여서 시스템을 구축하게 된다면 기존 웹 게시판 정보수집 방법들에서 낭비되는 시간과 비용의 절감이라는 목적에서 비용 적 측면이 의미를 잃는다[4].

3.2 전체 시스템 구조

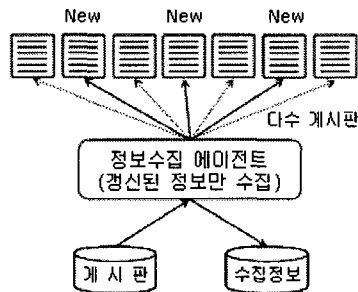
본 논문에서 제안하는 시스템 구조는 [그림 1]과 같으며 각 부분의 작동 프로세스는 아래와 같다.



▶▶ 그림 1. 전체 시스템 구조

1) 정보수집 에이전트

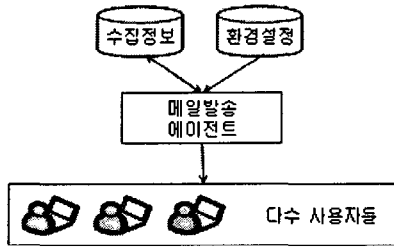
정보수집 에이전트는 주기적으로 사용자가 등록한 "목표 게시판 등록정보"를 읽은 후 그 등록정보에 따라서 등록된 게시판들을 접속하여 갱신된 정보를 추출하여 로컬데이터베이스에 저장한다.



▶▶ 그림 2. 정보수집 에이전트의 작동 프로세스

2) 메일발송 에이전트

메일발송 에이전트는 주기적으로 사용자가 등록한 "사용자 환경설정" 데이터베이스에서 사용자가 이미 등록한 "키워드" 와 "제약환경" 그리고 "메일주소 정보" 등을 읽은 후 그에 따른 쿼리를 "수집정보 데이터베이스"에 보내어 검색결과를 돌려받은 후 이를 사용자한테 메일로 발송한다.

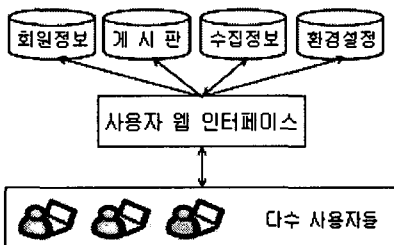


▶▶ 그림 3. 메일발송 에이전트의 작동 프로세스

3) 사용자 웹 인터페이스

사용자 웹 인터페이스는 다음과 같다.

- 사용자 등록 및 인증: 사용자는 웹 인터페이스에 접속하여 사용자 등록을 하며 이 정보는 "사용자 등록정보" 데이터 베이스에 저장되며 사용자 인증에 사용된다.
- 목표게시판 등록정보: 정보수집을 원하는 다수의 목표게시판의 정보를 "목표 게시판 등록정보" 데이터 베이스에 등록 및 수정 삭제를 한다.
- 사용자 환경설정: 미리 검색조건 및 키워드를 설정해서 저장해 놓고 검색 시에 키워드 입력과 검색조건 옵션을 선택할 필요가 없이 단 한번의 클릭으로 검색이 가능하게 할 수 있으며 이 검색결과를 자동메일수신 여부를 설정할 수 있다.
- 고급 검색: 등록된 게시판들에서 수집한 정보들에 대하여 키워드별, 시간대별, 게시판별, 결과물 출력개수 지정, 게시물 제목, 게시물 작성자, 게시물 본문 등 다양한 옵션으로 검색이 가능하며 검색한 결과를 등록시간대 순, 제목 순, 작성자 이름 순으로 자유롭게 정렬하여 열람이 가능하다.



▶▶ 그림 4. 사용자 웹 인터페이스의 작동 프로세스

IV. 결론

본 논문에서는 다중 웹 게시판 정보 모니터링 시스템이 여러 웹 게시판을 일정 시간간격으로 방문하여 최신 갱신정보를 확인하여 수집하고, 수집한 정보를 데이터베이스에 저장하고 사용자들은 일관된 인터페이스를 통하여 데이터베이스에 저장된 수집정보를 편리하게 검색하고 획득할 수 있는 "다중 웹 게시판 정보수집 및 관리시스템"을 개발하였다. "다중 웹 게시판 정보수집 및 관리시스템"은 원하는 정보를 찾기 위하여 일일이 여러 웹 게시판을 향해해야 하는 번거로움을 해소할 뿐 아니라, 단일 게시판검색 및 통합검색 그리고 게시물 정렬 등 강력한 관리 기능도 가능하며, 또한 수집한 정보를 로컬 데이터베이스에 저장하여 관리하므로 원격 웹 게시판이 사라졌다고 해도 이미 수집한 정보는 소장할 수 있다.

■ 참고 문헌 ■

- [1] 전진욱 "정보 수집 에이전트를 사용한 어린이 온라인 교육 정보 검색 시스템의 설계 및 구현", 대구가톨릭대학교 교육대학원 석사학위논문, pp.35, 2002.
- [2] Genesereth M. and Ketchpel S., "Software agents," *Comm. ACM*, Vol. 37, No.7, pp 48-53, 1994
- [3] 최중민 "에이전트의 개요와 연구방향", 정보과학회지, 제15권 제3호, 1997.
- [4] 최준권 "사용빈도 높은 웹 정보의 자동검색 시스템 개발", 경희대학교 대학원 전자공학과 석사학위논문, pp.7-8, 2002.