

이동 에이전트를 이용한 분산 정보수집 시스템

공용해* 최인석**
*순천향대학교 정보기술공학부
**홍성기능대학 전자계산기과
cis@kopo.or.kr

Distributed Information Gathering System using Mobile Agents

Yong Hae Kong* In Seok Choi**
*Div. of Information Technology Engineering, Soonchunhyang University
**Dept. of Computer, HongSung Polytechnic College

요 약

객체 컴포넌트 기반 웹 정보 저장/검색 시스템인 IHWA 시스템의 정보 수집 효율을 증대하기 위하여 이동 에이전트를 이용한 분산 정보수집 시스템을 구현하였다. 분산 정보수집 시스템은 이동 에이전트를 약간의 제약을 갖는 준 협약 사이트로 파견하여 XML 문서를 파싱하고, 정보를 추출하여 IHWA 시스템의 데이터베이스에 저장하도록 한다. 이동 에이전트는 준 협약 사이트에서 XML 파서를 활용하여 필요한 정보만을 수집하여 전송하므로 정보의 전송량을 최소화하여 네트워크의 부하를 줄일 수 있으며, 동시에 IHWA 시스템의 처리 부하를 크게 줄일 수 있다. 또한 향후 원격 사이트의 고유한 문서 특성에 적합한 정보 추출이 가능하도록 확장할 수 있다.

1. 서론

사용자가 정보 검색을 쉽게 할 수 있도록 개발된 IHWA (Information Harvest WArehouse) 시스템은 객체 컴포넌트 기반 웹 정보 저장/검색 시스템으로서 클라이언트의 요구에 의해 데이터베이스에 대한 쿼리를 수행하고, CORBA/IIOP를 통해 인터넷상에 존재하는 협약된 여러 IHWA 사이트간의 정보교환이 가능하여 일관성 있는 정보를 보유할 수 있어야 한다[1].

본 연구는 IHWA 시스템의 다양한 양질의 정보를 수집하기 위하여 약간의 제약을 갖는 준 협약 사이트에 에이전트를 파견하여 XML 문서를 파싱한 후에 정보를 수집하는 분산 정보수집 시스템을 구현하였다. XML 파서는 XML 문서를 읽고 문서의 내용과 구조를 알 수 있게

해주므로 XML 애플리케이션의 기반이 될 수 있으며, XML 문서를 해석하는 기능과 파싱을 통해 얻어진 각종 정보를 제공한다[2]. 정보를 전달하는 과정에서 SAX를 이용한다. SAX는 그 이벤트가 일어났을 때 제어권을 가지고 직접 XML 파서와 함께 상황을 처리하며 어떤 요소, 속성 등이 나타났는지 알려주는 구조 API이다[3].

원래의 호스트(Host)로 전송된 수집된 정보는 IHWA 시스템의 DB에 저장하고 사용자에게 보다 다양한 정보를 제공하게 된다. 이러한 정보 수집 방법은 정보를 수집한 후에 전송하므로 통신망과 호스트에 부하를 가중시키는 기존의 웹 스파이더 방식의 단점을 보완할 수 있다.

원격 사이트의 정보 수집 대상이 되는 XML 문서는 적절히 구성된 DTD (Document Type Definition)를 가지고 있는 것으로 가정하였다. DTD는 문서 유형에 사용된 요소를 선언하고 이 요소들 간의 관계를 기술하는 등의 역할을 한다. XML 파서는 간혹 sax를 활용하여 적절히 작성된

본 연구는 정보통신부의 ITRC 사업에 의하여 수행된 것임

DTD를 가진 XML 문서를 파싱함으로써 원하는 정보를 추출해 낼 수 있다.

본 연구에서 구현한 이동 에이전트를 이용한 분산 정보수집 시스템은 준 협약 사이트에 이동 에이전트를 파견하여 정보를 수집한 후에 IHWA 시스템에 전송함으로써 사용자에게 다양한 정보를 제공하게 한다. 이러한 정보 추출작업을 원격 사이트에서 실행하여 필요한 정보만을 전송하기 때문에 호스트와 네트워크의 부하를 줄일 수 있는 장점을 가지며 원격 사이트의 특성에 따라 적용될 수 있도록 확장이 가능하다. 이동 에이전트를 이용한 분산 정보수집 시스템이 추가된 IHWA 시스템의 전체적인 구성도는 그림 1과 같다.

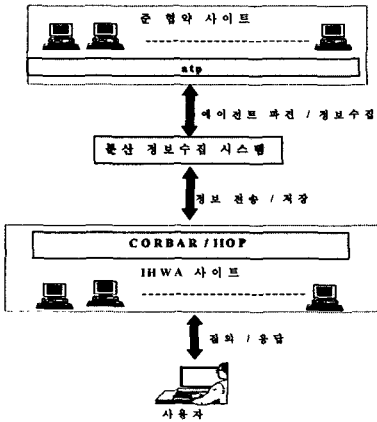


그림 1 분산 정보수집 시스템 구성도

2. 이동 에이전트 시스템

이동 에이전트란 사용자를 대신하여 이 기종 분산 환경의 네트워크에서 자율적으로 이동 및 반응하면서 필요한 정보를 수집하는 소프트웨어를 의미한다[4]. Java 기반 이동 에이전트 시스템은 Java 가상머신(Java Virtual Machine)상에서 구현되었으며 이 기종 분산 환경의 네트워크에서 이동이 원활한 이동 에이전트 시스템이다. Java 기반 이동 에이전트 시스템의 개발을 위하여 IBM사의 Tokyo Research Lab에서 Aglets Workbench를 개발하였다[5]. Aglets Workbench는 실행 환경과 함께 이동 에이전트 프레임워크를 Java 클래스 라이브러리 형태로 제공하여 에이전트 개발을 쉽게 하였다. Aglet은 Tahiti라 불리는 GUI 기반의 Java 응용프로그램이다[6]. 이동 에이전트에서는 웹서버를 두지 않고 Tahiti를 이용한다. 또한 이동 에이전트 객체를 원격 사이트로 보낼 때에도 Tahiti

를 이용한다.

3. 분산 정보수집 시스템의 구축 환경

IHWA 시스템은 기존의 객체 컴포넌트를 통합함으로써 새로운 응용프로그램을 작성할 수 있는 CBSE (Component Based Software Engineering) 기법을 바탕으로 개발되어진 웹 기반 정보 저장/검색 시스템이다. IHWA 시스템의 구조는 EJB (Enterprise Java-Beans)와 JCC (Java Commerce Client) 객체 컴포넌트를 기반으로 구현되어져 있다.

IHWA 시스템에는 Query Agent와 Rebuild Agent라 불리는 2개의 중요한 서버 측 컴포넌트가 있다. Query Agent는 클라이언트의 요구에 의해 데이터베이스에 대한 쿼리를 수행하고, 정보를 클라이언트들과 다른 시스템에 제공한다.

Rebuild Agent는 CORBA/ IOP를 통해 인터넷 상에 존재하는 여러 IHWA 사이트간의 정보교환작업을 수행하여 일관성 있는 정보를 유지하게 된다. 그러나 모든 인터넷 쇼핑물들이 상품 카탈로그 및 보유한 정보를 검색하고 저장하기 위하여 IHWA를 채택한다고 가정할 수는 없다. 그러므로 IHWA시스템은 웹 상에서 광범위한 검색 능력을 갖기 위하여 협약되지 않은 인터넷 쇼핑물에서도 상품정보를 수집할 수 있어야 한다.

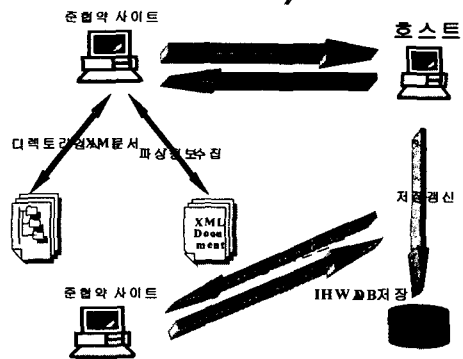


그림 2 분산 정보수집 시스템의 기능

본 연구에서는 보다 나은 시스템의 구성 및 정보의 다양성을 높이기 위하여 협약되지 않고 약간의 제약 조건을 갖는 준 협약 사이트에서 필요한 정보를 추출하여 IHWA시스템에 제공하는 이동 에이전트를 이용한 분산 정보수집 시스템을 구축하였다.

정보수집 이동 에이전트는 준 협약 사이트에서 표

준화 된 상품 정보만을 수집하는 역할을 담당한다.

상품정보 표준화는 각각의 상품 종류마다 DTD를 구성하게 하고, XML 데이터로 표현되도록 하여 문제를 해결하였다. 이동 에이전트는 준 협약 사이트에 파견되어 XML로 표현된 상품 카탈로그 문서를 찾아서 상품정보만을 추출한 후 이를 원래의 사이트에 전송하여 IHWA데이터베이스에 저장하게 한다. 그림 2는 정보수집 이동 에이전트시스템의 기능을 보여주고 있다.

3. XML 파서

데이터의 형식을 검증하기 위한 DTD와 이를 지원하는 도구들이 존재한다. 문서를 가시화하고 변환하기 위한 스타일시트(XSL)와 쿼리언어(XQL)가 지원된다. XML문서를 분석하여 각 태그의 의미를 파악하고 필요한 정보를 추출하는 일은 파서가 담당한다.

XML 문서를 읽고 문서의 내용과 구조를 알 수 있게 해주는 XML파서는 XML 애플리케이션의 기반이 될 수 있다. XML 애플리케이션은 XML 문서의 내용과 구조를 파악하고 적절히 처리하여 호스트에 전달해 주는 역할을 하는 모듈이다.

XML 파서의 기능은 첫째 XML 문서의 선언부와 프롤로그 정보를 이용해 문서를 해석하는 기능을 제공한다. XML 문서는 DTD(Document Type Definition)를 참조해 만든 유효한(Valid) 문서와 DTD 참조 없이 만든 잘 구성된 문서로 구분할 수 있는데 각 문서의 오류를 검사하는 기능을 제공한다. 또한 발견된 오류에 대한 정보와 수정방법에 대한 정보를 사용자에게 제공한다. 둘째, 문서의 파싱을 통해 얻어진 엘리먼트나 엔터티의 구성정보, 현재 파싱되고 있는 문서의 위치 및 파싱 상태 정보, 파싱이 끝난 후 재가공된 문서 정보(문서의 트리나 DTD 트리 등)를 제공하여 XML 애플리케이션에서 활용할 수 있는 XML 데이터 구조 제공한다. 애플리케이션에 전달하는 과정에서 때로 SAX(Simple API for XML) 또는 DOM(Document Object Model) 같은 표준 인터페이스를 사용한다.

SAX는 XML을 사용하는 웹 파일들을 해석할 수 있게 해주는 대표적인 이벤트 중심의 인터페이스이다. 프로그래머가 일어날 수 있는 이벤트를 설정해 놓으면, SAX는 그 이벤트가 일어났을 때 제어권을 가지고 직접 XML 파서와 함께 상황을 처리한다. SAX는 XML 문서를 앞에서 뒤로 읽어가면서 어떤 요소, 속성 등이 나타났는지

알려주는 구조 API이다. SAX 파서는 XML 문서를 정보의 흐름으로 바꾸는 도구이며 문서를 처음에서 끝까지 순차적으로 처리한다. 본 연구에는 SAX를 이용하였다.

4. 분산 정보 수집 시스템

준 협약된 사이트에 이동 에이전트를 파견하여 적절히 구성된 XML 문서로부터 정보를 수집하는 분산 정보수집 시스템을 구축하였다.

본 연구에서 구현한 분산 정보 수집 시스템은 디렉토리 리스팅, XML 파싱, 정보 수집 등의 작업을 원격 사이트에서 수행하도록 하였다. 그림 3은 분산 정보수집 시스템의 각 기능들을 상세하게 도식화한 것이다.

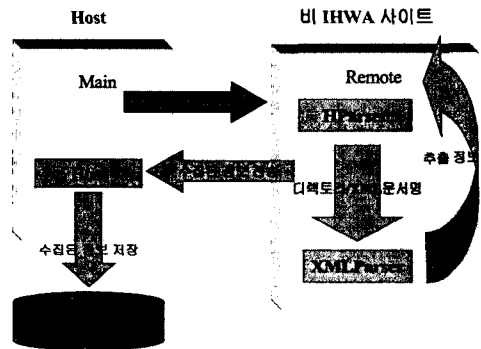


그림 3 분산 정보수집 시스템의 상세도

호스트는 Host 에이전트와 Remote 에이전트를 생성한다. Remote 에이전트는 준 협약 사이트에 파견된다. 파견된 Remote 에이전트는 Hparser를 이용하여 준 협약 사이트의 디렉토리를 조사하여 확장자가 XML인 문서 파일을 찾는다. XMLparser는 SAX를 활용하여 찾아진 XML 문서의 요소와 속성을 파악하여 SQL을 생성한다. 생성된 QSL은 Remote 에이전트에 의해 호스트에 전송되며 Host 에이전트는 IHWA 시스템의 데이터베이스에 저장한다. 모든 정보 수집 활동이 완료되면 Remote 에이전트는 수집된 정보를 전송하고, 호스트로 귀환 또는 제거하도록 하였다. 그림 4는 구현된 분산 정보 수집 시스템의 가상 코드(pseudo code)이다.

호스트 :

1. 에이전트 생성
2. Remote 에이전트 생성
3. 에이전트 사건발생 감지
4. 에이전트 행동 개시

5. 현재 호스트 URL의 확보
6. Remote 에이전트 파견
7. 전송된 정보를 IHWA 시스템 DB에 저장
8. 에이전트 제거

원격 사이트 :

1. Gatherer 생성
2. Hparser의 디렉토리 리스팅
3. XMLparser의 XML문서 파싱 // SAX 활용
4. 정보 수집
5. SQL 생성
6. 생성된 SQL 호스트 7로 이동(정보 전송)
7. 에이전트 제거

그림 4 이동 에이전트 가상코드

구축된 분산 정보수집 시스템은 이동 에이전트를 준 협약 사이트 사이트에 파견하여 정보 수집 작업을 수행하고 수집된 정보만을 가지고 원래의 호스트로 이동한다. 이러한 정보 수집은 정보 IHWA 시스템으로의 전송량을 최소화하여 네트워크의 부하를 줄일 수 있으며 IHEW 호스트의 부하를 크게 줄일 수 있었다.

5. 결론

웹 정보 저장/검색 시스템인 IHWA에 보다 다양한 정보를 제공하기 위하여 준 협약 사이트에서 정보를 수집해주는 이동에이전트를 이용한 분산 정보수집 시스템을 구현하였다. 이러한 분산 정보수집 시스템은 이동에이전트를 통해 약간의 제약을 갖는 준 협약 사이트에서 수집된 정보들은 분산된 IHWA 호스트간에 서로 공유될 수 있으므로 사용자는 어떠한 IHWA 호스트에 접속하더라도 보다 다양한 정보를 검색할 수 있게 된다.

본 연구에서 구현한 분산 정보수집 시스템은 에이전트를 생성하여 원격지의 준 협약 사이트에 파견한 후, 원격 호스트의 디렉토리 리스팅을 통하여 XML문서를 검색한다. 이동에이전트 시스템은 XML파서를 이용하여 검색된 XML문서에서 필요한 상품정보만을 추출하여 정보를 수집한다. 수집된 상품정보는 원래의 호스트로 전송하여 IHWA 시스템의 DB에 저장하도록 한다. 파견된 이동에이전트는 준 협약 사이트에서의 정보수집 활동이 완료되면 호스트로 귀환시키거나 제거하였다. 분산환경에서 폭넓게 분포되어 있는 이 기종간의 이동이 자유로운 이동 에이전트에 의한 정보수집은 웹 스파이더 방식의 중앙 정보 수집의 단점인 과도한 통신 부하와 처리 부하를 완화

할 수 있었다. 또한 이러한 정보수집 이동에이전트는 원격 사이트의 특수한 형태의 정보 수집에 적합하도록 확장 가능하므로 향후 정보 수집의 질적 향상을 기대할 수 있다.

참고 문헌

- [1] Incheon Paik, Tongwon Han, "Design and Implementation of electronic commerce Search Engine Component", Proceedings of AoM/IAoM 17, August. 1999.
- [2] http://xml.t2000.co.kr/products/intro_parser.html.
- [3] <http://maso.zdnet.co.kr/maso>
- [4] 석황희, 김인철 "이동 에이전트의 개념과 응용", 한국멀티미디어학회, 제3권, 제2호, pp29-39, 1999.
- [5] <http://www.tri.ibm.co.jp/aglets>.
- [6] Danny B. Lange, "Programming and Deploying Java Mobile Agents with Aglets", ADDISON WESLEY, 1998.