

다중 에이전트 기반 웹 웨어하우징*

김현희^U 박승수
이화여자대학교 컴퓨터학과
{981COG03, sspark}@mm.ewha.ac.kr

Web Warehousing based on Multi-Agent

Hyon-Hee Kim^U Seoung-Soo Park
Dept. of Computer Science & Engineering, Ewha Womans University

요 약

본 연구에서는 기존의 데이터 웨어하우징 기술과 웹 기반 기술을 통합한 웹 웨어하우징 기법에 다중 에이전트 패러다임을 적용하여 다중 에이전트 기반 웹 웨어하우징 시스템을 설계, 구현하였다. 시스템은 정보 검색 에이전트, 정보 통합 에이전트, 웹하우스 구축 에이전트로 구성된다. 정보 검색 에이전트는 여러 종류의 웹 자원을 수집한다. 정보 통합 에이전트는 정보 검색 에이전트에 의해 수집된 이형질적인 데이터를 일정한 형식으로 변환한다. 웹하우스 에이전트는 생성된 데이터를 사용하여 웹하우스를 구축하고 관리한다. 웹 데이터를 통합하기 위해 새로운 데이터 모델을 제안하였다. 의미를 갖는 지능적 객체를 생성하기 위해 여러 종류의 추론 에이전트들이 추론작업을 수행하고, 이들은 블랙보드 시스템을 통하여 작업을 통합한다. 본 시스템은 의미 정보 데이터를 사용하므로 웹 정보의 의미적 검색과 정보추출이 가능하다.

1. 서론

웹 웨어하우징이란, 기존의 데이터 웨어하우징 기법과 웹 기반 기술을 통합한 새로운 형태의 지식 관리 방법을 뜻한다[1]. 월드 와이드 웹 자체가 거대한 정보 자원이 됨에 따라서 웹의 정보들을 사용자의 필요에 따라 관리하는 것이 중요한 이슈로 자리잡고 있다. 기존의 데이터 웨어하우징 기법은 이미 고정된 스키마를 갖는 데이터베이스들을 주제 중심으로 통합하는데 그 초점이 있기 때문에 웹 자원을 통합하는 데는 어려움이 있다. 이는 웹 데이터가 다음과 같은 특징들을 갖기 때문에 발생한다[2]. 데이터의 구조가 불규칙적이고 암시적이다. 또한 스키마가 고정되어 있지 않고, 빠르게 변화하며 대개의 경우 매우 크다. 따라서 이러한 문제점들을 해결하기 위해서는 기존의 관계형 데이터 모델과는 다른 새로운 형태의 모델이 필요로 되어지며, 이것은 반구조화된 데이터 모델로써 표현이 가능하다.

최근, 컴퓨팅 환경이 네트워크 중심으로 변화함에 따라 폭발적으로 증가하는 다양한 형태의 대용량의 정보를 다루기 위해서 다중 에이전트 시스템에 대한 관심이 증대되고 있다. 다중 에이전트 구조에서 소프트웨어

에이전트는 이들로 이루어진 더 큰 사회의 구성원으로서 역할을 수행한다. 이러한 개념은 Marvin Minsky의 SOM(Society of Mind)이론[3]에 기반한 것으로 민스키에 따르면 마음이란 매우 작은 많은 프로세스들로 구성되는데 아주 간단한 일들을 하는 이러한 에이전트들이 특별한 방법으로 조합될 때 마음이 생성된다고 하였다.

본 연구에서는 웹 웨어하우징을 위한 전 단계를 에이전트들이 담당하여 전체 시스템을 통합하는 다중 에이전트 기반 웹 웨어하우징 시스템을 제안하였다. 웹 웨어하우징은 크게 세 단계로 나뉜다. 첫 번째 단계는 에이전트들이 웹 공간에 퍼져있는 정보를 모으는 과정이다. 이 단계를 "정보 수집 단계"라고 한다. 두 번째 단계는 "정보 통합" 단계이다. 이 단계에서 에이전트들은 여러 가지 다양한 형태를 갖는 데이터들을 변형하고, 통합하는 과정을 거친다. 마지막 단계는 "웹하우스 구축" 단계이다. 이 단계에서 에이전트들은 통합된 정보를 저장하기 위한 웹하우스를 구축한다.

웹 데이터를 위한 웹하우스를 구축하기 위해서 새로운 데이터 모델이 필요하다. 기존의 관계형 데이터 모델은 웹 데이터를 표현하는데 부적절하며, 의미 정보를 포함하지 않는다. 따라서 본 논문에서는 의미 정보를 포함하며, 추론이 가능한 지능적인 객체로서 데이터를 표현하는 새로운 모델을 제시한다.

* 본 연구는 교육부 BK 사업의 지원으로 수행되었음

본 논문의 구성은 다음과 같다. 2장에서는 정보 통합 시스템의 두 가지 접근 방식을 살펴보고, 3장은 전체 시스템 구조를 설명한다. 4장에서는 새로운 데이터 모델을 제안하고 5장에서 시스템 구현을 살펴본다. 마지막으로 6장은 결론 및 향후 연구과제를 제시한다.

2. 관련연구

이형질적인 정보 자원들을 통합하기 위해 반구조화된(semi-structured) 데이터 모델을 사용한 예는 스탠포드 대학의 TSIMMIS Project[4]를 들 수 있다. 이 시스템은 다음과 같은 컴포넌트들로 구성된다. 먼저, 데이터 모델로서 OEM이라고 하는 "lightweight" object model을 제안한다. 이 모델은 self-describing이 가능하고 전통적인 객체지향 시스템의 모든 구조를 사용하는 것이 가능하다. 다음으로 Mediators가 있다. Mediators는 다양한 정보 자원을 통합하는 기능을 제공하며 Mediator Specification Language(MSL)에 의해 지정된다. Wrappers는 사용자의 질의를 도메인에 적절한 형태로 변환하는 역할을 한다. Wrapper역시 Wrapper Specification Language(WSL)로 지정된다. 마지막 요소는 이들을 만들어내는 Wrapper/Mediator generators다. 이 시스템은 이형질적인 정보 자원을 통합하기 위해 반구조적인 데이터 모델을 사용하는 대표적인 통합 시스템이다.

유사한 작업을 수행하는 에이전트 기반 접근 방법에는 MCC 연구소의 Infosleuth 프로젝트[5]를 들 수 있다. 이 연구는 인터넷의 정보를 수집하고 분석하기 위한 에이전트 기반 구조를 제공한다. 시스템은 몇 가지 종류의 에이전트들의 클래스로 구성된다. 정보 추출 에이전트는 데이터베이스, html/xml/text 파일, 이미지 자원 등으로부터 원하는 정보를 추출한다. 정보 분석 에이전트는 여러 종류의 질의를 통하여 정보를 분석한다. 작업 수행 에이전트는 작업을 수행하기 위한 전략을 세운다. 브로커 에이전트는 적절한 작업을 수행할 수 있는 에이전트들을 찾아준다. 이 시스템은 자원의 의미적 통합을 위해 도메인 온톨로지를 사용한다. 온톨로지와 이들을 표현하는 메타 모델로서 정보 통합이 이루어진다. 이 연구는 이형질적인 정보를 의미적으로 통합하기 위해 다중 에이전트 기반 기술을 사용한 예이다.

3. 전체 시스템 구조

전반적인 시스템 구조는 다음 그림 1과 같다. 여러 종류의 데이터베이스, html/xml 문서, 이 밖에 웹을 통해 접근 가능한 모든 종류의 데이터들이 정보자원이 된다. 먼저, 정보 수집 단계를 위해 정보 검색 에이전트들이 질의를 수행하여 정보를 수집한다. 질의 결과 수집된 데이터들은 구조화되지 않고 자원에 의존적인 데이터 형태를 갖는다. 다음으로, 이 정보들을 통합하는

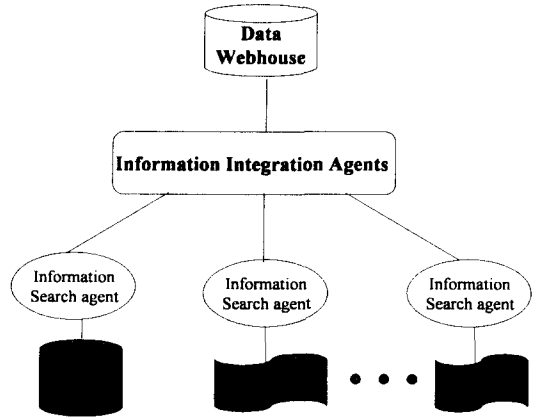


그림 1. 전체 시스템 구조

작업은 정보 통합 에이전트들이 담당한다. 정보 통합 에이전트들은 정보 검색 에이전트들로부터 받은 정보를 새로운 데이터 모델로 변환한다. 데이터 모델에 의미를 부여하기 위해서 여러 종류의 추론 에이전트들이 협력한다. 추론 에이전트의 종류로는 사례 기반 추론 에이전트, 규칙 기반 추론 에이전트, 비단조 추론 에이전트 등이 있다. 마지막으로 웹하우스 구축의 전 단계를 담당하는 웹하우스 에이전트들이 있다. 웹하우스 에이전트는 생성된 데이터를 웹하우스에 저장하고, 웹하우스를 유지 및 관리한다.

4. 데이터 모델

제안하는 데이터 모델은 OEM(Object Exchange Model)[6]을 기반으로 한다. OEM 모델은 반구조화된 데이터 모델을 표현하기에 유용한 데이터 모델이다. 이 모델은 에지에 대한 레이블로 그래프를 구성하고 모든 엔터티들은 객체로 표현된다. OEM은 관계형 데이터, 객체형 데이터, 계층형 데이터, 그래프 데이터 모두를 쉽게 표현한다. 특히 그래프로 표현이 가능하므로 웹 정보의 하이퍼텍스트 특성을 잘 표현할 수 있다.

OEM은 그 자체로 의미표현이 불가능하다. 기존의 에이전트 기반 접근 방식[7]에서 정보에 의미를 부여하기 위해 사용했던 방식은 다음과 같다. 특정 도메인에 대한 어휘라고 할 수 있는 온톨로지와 이를 표현할 수 있는 KIF(Knowledge Interchange Format)로 구성된다. 온톨로지는 일반적인 응용프로그램에 적당한 단어들이나 사전으로, 사전은 주어진 영역에 대해 여러 개의 온톨로지를 포함할 수 있다. KIF는 일차논리의 전위 표현 방식으로 간단한 데이터의 표현을 지원하며, 논리적 요소, 인덱싱, 비단조 추론 등이 가능하다.

따라서, 제안하는 모델은 기존의 OEM에 의미 요소를 결합한 지능형 객체다.

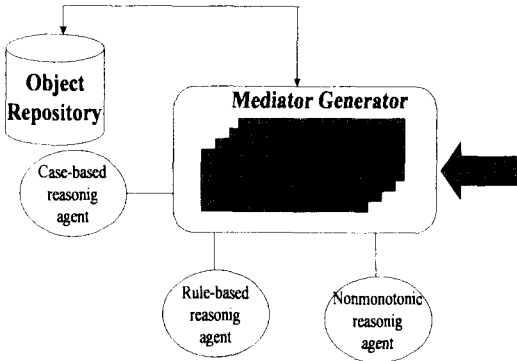


그림 2. Information Integration Architecture

정보 통합을 위한 구조는 그림 2와 같다. 정보 통합 에이전트는 블랙보드 시스템으로 구성된다. 의미를 갖지 않는 OEM 객체가 입력되면, 여러 종류의 추론 에이전트들이 의미를 부여한다. 블랙보드 시스템[8]은 각 추론 프로세스들이 이론을 발전시키기 위해 필요한 작업 영역을 제공한다. 새로운 사실이 생성되면 합당한 규칙이 적용되고, 추론 에이전트들은 이 사실과 규칙들을 개별적으로 블랙보드에 포스팅하여 서로 협동한다.

규칙 기반 추론 에이전트는 자료 형태의 불일치를 해결하기 위한 규칙들을 기반으로 추론한다. 사례기반 에이전트는 이전에 생성되었던 객체들에 대한 규칙을 사례 데이터베이스에 저장한다. 유사한 경우가 입력되면, OEM 객체에 의미를 주는 방식으로 이전의 사례를 택한다. 비단조 추론 에이전트는 의미 추출을 위한 단계에서 사용된다.

5. 시스템 구현

데이터 모델을 구현하기 위해서 XML기반 기술[8]을 사용한다. XML은 OEM 표현이 가능하며 다음과 같은 장점을 갖는다. 첫째, 웹을 통해서 접근이 가능하다. 둘째, 스키마가 고정되어 있지 않기 때문에 새로운 데이터 모델을 표현하는데 적절하다. 셋째, 문서의 논리 구조를 가지고 있기 때문에 의미 표현을 용이하게 한다. 새로운 데이터 모델을 위한 추론 에이전트들은 CLIPS로 구현된다. CLIPS는 LISP으로 구현된 전문가 시스템 툴이다. 따라서 에이전트들의 추론 구조를 적절히 표현할 수 있다. 순수 자바 코드로 구현하기 위해서 CLIPS의 자바 버전인 JESS를 사용한다. 블랙보드 시스템의 공유 메모리를 위해서 관계형 데이터 베이스를 사용한다. 데이터베이스와 에이전트들은 JDBC를 사용하여 연결된다. 각 추론 에이전트들은 관계형 질의문을 사용하여 다른 에이전트들과 의사 교환한다. 전체 시스템이 순수 자바 코드로 구현되므로 네트워크 컴퓨팅이 가능하다. 또한, 별도의 클라이언트 프로그램이 필요 없고, 웹을 통한 정보 접근이 가능하다.

6. 결론 및 향후 연구

본 연구에서는 웹 스페이스상의 정보를 검색하고 데이터 구조를 통합하여 웹하우스를 구축하는 웹 웨어하우스에 다중 에이전트 패러다임을 적용하였다. 시스템은 정보 검색 에이전트, 정보 통합 에이전트, 웹하우스 에이전트로 구성된다. 정보 검색 에이전트는 여러 종류의 데이터베이스, html/xml등의 웹 데이터를 수집한다. 정보 통합 에이전트는 각기 다른 형태로 모아진 데이터를 새로운 통합 모델로 변환하며 이 과정에서 데이터에 의미를 부여한다. 웹하우스 에이전트는 생성된 데이터를 웹하우스에 저장하고 관리한다. 또한, 의미를 갖는 새로운 데이터 모델을 제안하였다. 이 모델은 기존의 반구조화된 자료를 표현하기 위해 사용되는 OEM을 기반으로 하며, 추론 에이전트들의 작업을 통하여 의미를 갖는 새로운 지능형 객체로 재생성된다. 의미를 갖는 새로운 데이터 모델은 이후 웹하우스 분석 단계에서 웹 자원의 의미 정보 추출을 가능하게 하며, 정보 검색 단계에서도 적용되어 의미 정보 검색을 가능하게 할 것이다.

향후 연구로는 웹하우스 분석 단계를 들 수 있다. 구축된 웹하우스를 데이터 마이닝 기법을 적용하여 분석하면, 유용한 지식의 발견이 가능해질 것이다. 여러 종류의 데이터 마이닝 알고리즘 중에서 데이터 모델에 적절한 알고리즘을 선정하여 데이터를 분석하는 작업을 진행 중이다.

참고 문헌

- [1] R. Mattison., "Web Warehousing and Knowledge Management", McGraw-Hill, 1999.
- [2] S. Abiteboul., "Querying Semi-Structured Data", in Proc. Intl. Conference on Database Theory, Delphi, Greece, January, 1997.
- [3] M. Minsky., "The Society of Mind", Simon and Schuster, New York, 1985.
- [4] H. Garcia-Molina, et al., "The TSIMMIS Project: Integration of Heterogeneous Information Sources", In Proc. IPSJ Conference, pp. 7-18, October, 1994.
- [5] F. Jerry., "Agent-Based Semantic Interoperability in Infosleuth", <http://www.mcc.com/projects/infosleuth>
- [6] M. R. Genesereth., S. P. Katchpel., "Software agents", CACM, Vol. 37, No. 7, pp. 48-53, 1994.
- [7] D. Rieken., "M: Anarchitecture of Integrated Agents", CACM, Vol. 37, No. 7, pp. 1994
- [8] E. Harold., "XML: Extensible Markup Language", IDG Books Worldwide, September, 1998.