

멀티 에이전트 기반 비교 쇼핑 시스템 설계

신주리⁰ 한상훈 이진명
충북대학교 컴퓨터학과, 첨단정보기술 연구센터
edutopia@aicore.chungbuk.ac.kr

Design of a Multiagent-based Comparative Shopping System

Ju-Ri Shin⁰ Sang-Hoon Han Keon-Myung Lee
Dept. of Computer Science, Chungbuk national University and AITrc

요 약

이 논문에서는 보다 효과적이고 편리한 서비스를 제공할 수 있는 전자상거래를 위한 다중 에이전트 기반의 확장된 비교 쇼핑 시스템을 제안한다. 이 시스템은 웹 크롤링(web crawling)을 통해 비교 쇼핑 시스템의 대상이 되는 웹 사이트들의 페이지 추출 정보를 입수한다. 각 쇼핑 사이트에서의 정보 추출을 위한 중심이 되는 랩퍼(wrapper) 기술은 먼저 정보가 있는 페이지를 가려내고, 정보가 있다고 판명되는 페이지들에서 상품 정보의 위치 즉 반복되는 패턴(pattern)을 추출하여 필요한 상품 기술 단위 정보를 뽑아내는 학습 알고리즘이며, 각 사이트에 맞게 만들어진 랩퍼 에이전트(wrapper agent)에 대해 유효성을 검사하는 방법론을 제시한다. 또한, 학습 시 필요한 지식(knowledge)으로서의 디렉토리(directory) 구성은 미리 만들어진 표준 카테고리(category)와 용어(terminology) 존재 하에 제한적이나마 새로운 디렉토리 요소에 대해 자동으로 확장할 수 있는 방법론을 제안한다.

1. 서론

인터넷 사용의 보편화와 더불어 인터넷 기반의 전자상거래를 통한 경제 활동이 급격히 확산됨에 따라 전자상거래를 위한 정보 제공자들의 노력 또한 활발해졌다. 국내의 인터넷 쇼핑 몰(shopping mall)을 통한 정보 제공자들도 나날이 늘고 있는 추세이다. 이로 인해 인터넷에서 상품을 구매하려는 구매자들은 상품을 구매하기 위해서는 각 사이트에서 상품의 검색과 나름대로의 검색한 결과에 대해 비교 종합하여 상품을 구입한다. 이러한 시스템에서 사용자의 직접적 정보 처리 과정은 전자상거래의 시간적 제약이 될 수 있다. 또한 컴퓨터를 통한 전자상거래는 공간에 대한 제약이 되고 있다. 전자상거래의 규모와 이용자의 폭발적인 증가로 말미암아 좀 더 지능적이고 편리한 전자상거래를 요구에 부응하여 현재 여러 비교 쇼핑시스템이 제안되었다. 현재 서비스중인 비교 쇼핑 시스템들은 쇼핑 물들을 대상으로 사용자가 원하는 상품의 정보 검색을 하여 그 결과를 비교 종합하여 상품결과리스트를 나타내어 최적의 상품을 구입하도록 하는 형태이다. 이 논문에서는 여기서 좀 더 나아가 다중 에이전트기법을 응용하고, 유무선 인터넷을 통합하여 비교 쇼핑 검색을 지원함으로써 기존의 전자상거래에서의 공간적, 시간적 제약을 완화할 수 있는 확장성이 큰 다중 에이전트 기반 비교 쇼핑 시스템을 제안한다. 이제까지 비교 쇼핑 시스템에서 이슈가 되어 오던 것은 정보를 제공하는 쇼핑 몰에서의 사용자의 정보 입력방식과 상품 검색 결과를 나타내는 웹 페이지의 표현방식이 각 쇼핑 몰마다 모두 다르므로, 각 사이트에 표현

된 정보 표현 형식을 학습하여 필요한 정보를 추출하는 랩퍼 에이전트(wrapper agent) 생성의 자동화이다. 다행히 각각의 쇼핑 물들은 독특한 출력형식을 가지기는 하지만, 여러 상품 검색 결과 페이지는 모두 같은 패턴을 가진다. 이 논문에서 제안한 시스템에서는 이에 대해 각 쇼핑 몰에 대한 검색 질의 결과 페이지에 대한 동일한 패턴학습과 정보 추출 방법을 제시하며, 정보의 처리과정에 있어 랩퍼 에이전트의 유효성 여부와 표준 온톨로지 생성과 그 확장 방법에 대해 많은 정보와 이에 대한 처리과정을 다중 에이전트 기법을 이용하여 효율적인 시스템을 구축한다. 비교 쇼핑을 하는데 있어 시간과 공간의 제약을 받지 않고 쇼핑을 하며, 쇼핑 몰 및 물품 판매 웹 사이트가 비교 쇼핑 시스템에 대해서 비협조적인 것을 전제하여 시스템이 동작하며, 비교 쇼핑 대상 물품이나 비교 쇼핑 대상 쇼핑 물 및 물품 판매 웹 사이트를 추가 확장할 수 있도록 한다. 비교 쇼핑 시스템에 새로운 쇼핑 물들을 쉽게 추가시킬 수 있도록 하고 기존 쇼핑 물의 구성 등의 변화가 일어나도 이를 검증하여 대응할 수 있도록 시스템을 설계하였다.

이 논문의 구성은 2절에서 제안한 다중 에이전트 기반 비교 쇼핑 시스템의 구조를 소개하고 3 절에서는 시스템의 구현 내용과 기술을 설명하고 마지막으로 4 절에서는 이 논문의 결론에 대해 기술한다.

2. 시스템 구조와 특성

2.1 사용자 인터페이스 에이전트(User Interface Agent)

사용자 인터페이스 에이전트는 웹 브라우저나 무선 인터넷 단말기로부터 비교 쇼핑할 물품에 대한 사양정보

이 논문은 첨단 정보기술 연구센터(AITrc)를 통해서 과학재단의 지원을 받은 것이다.

및 검색조건 등을 입력 받아서 비교 쇼핑 에이전트에게 전달하는 역할을 한다. 비교 쇼핑 결과는 전자우편이나 단문 메시지 전송(SMS)을 통해서도 전달하는 기능을 가진다.

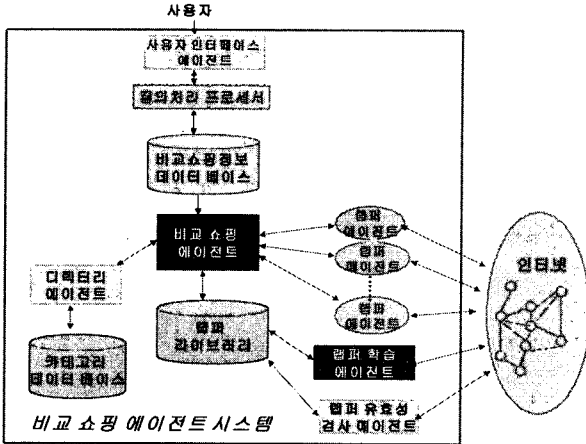


그림 1. 멀티 에이전트 기반 비교 쇼핑 시스템 구조

2.2 비교 쇼핑 에이전트(Comparative Shopping Agent)

비교 쇼핑 에이전트는 비교 쇼핑 정보 데이터 베이스를 구축하기 위해 디렉토리 에이전트에게 검색할 제품 카테고리라 접속해 보아야할 쇼핑 물 및 판매 웹 사이트에 대한 정보와 이들로 부터 정보를 얻어서, 이를 이용하여 상품 정보를 수집할 웹퍼 에이전트를 웹 사이트별로 하나씩 활성화 시킨다. 웹퍼 에이전트가 생성한 결과는 비교 쇼핑 에이전트가 비교 쇼핑 정보 데이터 베이스에 체계적으로 저장한다.

2.3 웹퍼 에이전트(Wrapper Agent)

에이전트는 비교 쇼핑 에이전트로부터 정보를 수집할 정보를 받은 다음, 일반 웹 브라우저처럼 HTTP (Hypertext transfer protocol)를 이용하여 쇼핑 물 및 판매 웹 사이트에 접속하여 각 사이트에 대해서 페이지 종류별에 따른 정보 저장 형태 규칙을 알아내는 것으로 결국엔 비교 쇼핑 대상 물품에 대한 정보를 수집하고자 한다. 쇼핑 물과 판매 웹 사이트의 특성이 서로 다르기 때문에 웹퍼 에이전트는 사이트별 특성에 맞게 구현된다. 활성화된 웹퍼 에이전트에서는 자신에게 할당된 사이트에 접속하여 정보를 수집하여 그 결과를 다시 비교 쇼핑 에이전트에게 보낸다.

2.4 웹퍼 학습 에이전트(Wrapper Learning Agent)

웹퍼 학습 에이전트는 주어진 URL의 모든 페이지를 입수하여 정보를 포함한 페이지를 구별하고 해당 페이지로부터 제품 기술단위(Product Description Unit)를 추출하는 규칙을 학습하여 이 정보를 웹퍼 라이브러리(Wrapper Library)에 저장하는 역할을 한다.

2.5 웹퍼 라이브러리(Wrapper Library)

웹퍼 학습 에이전트에 의해서 구축되고 웹퍼 라이브러리는 쇼핑 물에 대한 정보를 포함한 페이지의 URI와 해당 페이지의 상품기술단위(PDU) 형태 상품 카테고리등의 저장 정보를 가지고 있다.

2.6 디렉토리 에이전트(Directory Agent)

디렉토리 데이터베이스에는 표준 제품 카테고리(category)

와 표준 용어(terminology)를 가지고 있다. 웹퍼 에이전트에서 받아들인 상품 정보에 대해 비교 쇼핑 정보 데이터베이스에 저장하기 앞서 표준 카테고리라 표준 용어에 맞게 매핑(mapping)시켜 주는 역할을 하며, 용어와 카테고리 자동 확장 기능을 담당한다

2.7 카테고리 데이터베이스(Category Database)

카테고리 데이터 베이스에는 온톨로지에 관한 것과 제품 카테고리 정보 즉 카테고리라 제품명등과 쇼핑 물에 대한 URL 및 쇼핑 물 정보가 저장되어 있다.

2.8 비교 쇼핑 정보 데이터 베이스

(Comparative Shopping Information Database)

비교 쇼핑 정보 데이터 베이스에는 각 쇼핑사이트 별로 카테고리, 제품명, 가격, URL 등과 같은 상품 기술 단위에 대한 데이터가 저장되어 있다.

3. 시스템 구현 내용과 기술

3.1 웹 크롤링(Web Crawling)

웹퍼 학습 에이전트는 보다 정교한 웹 크롤링 방법을 사용하여 비교 쇼핑 시스템의 대상이 되는 쇼핑 사이트들의 정보를 가져온다. 페이지들에서 모든 앵커(anchor)를 추출한다. 그리고 앵커 정보를 함께 추출하며, URL의 list를 만든다. 이 방법은 기존의 URL list에서 새로운 URL이 들어오면 체크를 하여 중복되는 것을 막으며, 또 하나의 URL에 대해 여러 가지로 표현된 것에 대해서도 알아 낼 수 있다. 이렇게 쇼핑 사이트들에 대해서 앵커정보를 추출하여 카테고리 검색을 하게 된다. 즉 사이트별 데이터 베이스를 구축한다.

3.2 상품기술단위(Product Description Unit) 추출

쇼핑 물 등의 사이트로부터 페이지를 입수하면 다음의 문제는 페이지들에 대한 정보의 존재 여부를 검사 한다. 앵커의 텍스트 정보, 반복 패턴 유무 등을 이용한 키워드 검색을 통하여 상품정보가 있다고 판단된 페이지에 대해서는 상품기술 단위 정보가 어떻게 상품 정보를 나타내는지 추출해야 한다. 시스템 구현을 위해 수십 개의 쇼핑 사이트 분석 결과, 상품 정보를 포함한 상품 기술 단위(PDU)는 대부분 테이블(table) 형식을 취하고 있었다. 테이블 형식은 다음 그림과 같이 크게 3가지로 나눌 수 있다.

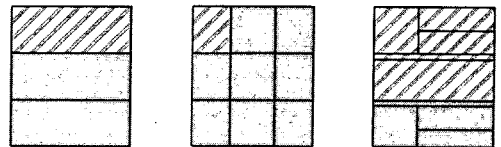


그림 2. 테이블 형태의 상품 기술 단위 유형

PDU 1은 레코드 하나가 한 상품인 경우이고, PDU 2는 테이블이 셀(cell) 단위로 정보가 표현된 경우이며, PDU 3는 레코드가 <tr> 태그간 형태가 다른 경우로 복잡구조를 갖는 경우이다. 테이블 형태는 먼저 아주 기초적인 형태로서 어트리뷰트(attribute)를 포함하는 테이블로서 표준 어트리뷰트 온톨로지라 확인하여 PDU를 추출할 수 있고, Tr 간의 단위로 PDU를 표현한 경우에는 각 상품 기술 단위는 일정한 반복 패턴을 가지므로, 휴리스틱(heuristic)과 정규식(regular expression)을 적용하여 PDU를 추출한다.

의미 있는 페이지 중 상품기술단위가 테이블 형식을 갖지 않는 경우에는 <HR>, LIST,

 를 이용하여 PDU를 구별하는 경우가 있다. 이 경우에는
 태그(tag)에 의해 시각적으로 한 줄로 표현되는 HTML 텍스트를 정규식(regular expression)으로 추상화 하여 변환하여 각각 독특한 형태를 갖춘 정규식에 고유 번호를 부여하여 이들 번호가 반복되는 패턴(pattern)을 추출하여 이 정보를 통해 PDU를 추출한다. 이와 유사한 기법은 Shop bot 등에서 이미 이용된 것이다.[3]

3.3 랩퍼 유효성 검사(Wrapper Validity Test)

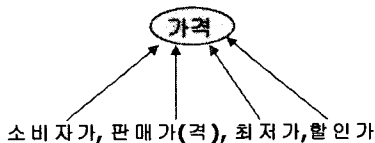
랩퍼의 유효성 검사는 랩퍼가 웹 사이트의 독특한 구조에 맞게 학습되어 생성되므로 그 웹 사이트의 표현 형식이 변하면 필요한 정보를 추출 할 수 가 없다. 따라서 제안한 방법에서는 두 단계를 거쳐서 특정 쇼핑 물에 대해 추출한 랩퍼 라이브러리의 내용을 검사한다. 일차 검사는 기존 페이지의 유무 및 새로운 페이지의 유무 검사이다. 기존 페이지가 없어진 경우에는 정보를 삭제하고, 새로운 페이지가 생성되었다면 랩퍼 라이브러리에 특정 쇼핑 물에 대한 페이지정보를 생성한다. 그 사이트 내 페이지에서의 큰 변화가 없는 경우에 한하여 이차 검사로 기존 페이지 변화를 검색하는 것으로 키워드(keyword) 검사를 한다 기구축의 PDU 정보와 비교하는 것이다. 일 이차 랩퍼 유효성 검사에 대해 랩퍼 에이전트가 특정 사이트에 대해서 유효하지 않다고 판명되면 그 사이트에 대해서 랩퍼 에이전트를 다시 생성하도록 한다.

3.4 디렉토리(Directory) 구축

각 사이트별 특성에 맞게 상품기술단위(PDU)정보를 추출한 뒤, 비교 쇼핑 정보 데이터 베이스에 저장되기 위해서는 그 추출한 상품에 대한 정보가 어떤 정보를 제공하는 지에 대한 인식이 필요하므로, 각 사이트에서 제공하는 용어(terminology)에 대해서 그 용어가 어떤 의미를 가지는 지를 알아야 한다. 이 논문에서 제안한 디렉토리 에이전트를 거쳐 표준 용어와 표준 카테고리에 맞게 매핑(mapping)된 후 비교 쇼핑 정보 데이터 베이스에 저장된다.

표준 용어와 카테고리는 미리 구축되어 있는 것으로서 여기서는 제한적인, 새로운 용어와 카테고리에 추가되는 새로운 상품에 대해서 자동적으로 디렉토리를 확장할 수 있는 방법을 제안한다.

3.4.1 용어(terminology) 확장 방법

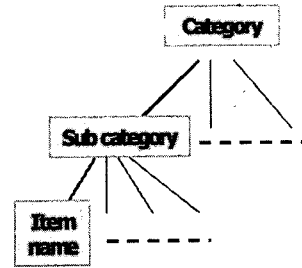


예를 들어, 현재 디렉토리는 표준 용어 “가격”에 대해서 소비자가, 판매가(격), 최저가, 할인가와 같은 용어에 대해 가격으로 매핑(mapping)이 이루어지도록 구성되었다고 하자. 새로운 쇼핑 물 사이트에서 가격에 대해 “회원가”와 같이 가격에 대해 새로운 용어가 나타나면, 해당 쇼핑 물의 상품 정보들을 상품 정보 데이터 베이스에서 검색해서 비교하여 그 용어의 의미를 알아 낸 후, 새로운 용어를 표준 용어(terminology)에 추가 하도록 한다.

3.4.2 카테고리(category) 확장 방법

먼저 표준 카테고리가 구성되어 있는 전제 하에서, 랩퍼가 상품 정보 수집 시 비교 쇼핑 시스템의 카테고리에 따라 제품 정보를 데이터 베이스에 저장하기 위해서는 첫번째로 앵커 텍

스트(anchor text)를 이용하거나 또는 상품 기술 단위(PDU) 정보를 이용하여 상품정보 데이터 베이스에서 대응하는 PDU 정보를 검색하여 카테고리를 구성한다. 새로운 카테고리의 출현 시는 디폴트 카테고리(default category)를 구성하여 오퍼레이터(operator)가 지정해주도록 한다. 마지막으로 새로운 아이템 이름(item name)이 발견되었을 경우에는 카테고리 정보를 이용하여 카테고리에 추가하도록 한다.



4. 결론

이 논문에서는 다중 에이전트 기반의 비교 쇼핑 시스템을 제안하였다. 비교 쇼핑 시스템 데이터 베이스를 구축하기 위해서 쇼핑 물 웹 사이트들의 페이지 추출 정보를 입수하여 각 사이트별 상품 정보 추출을 위한 학습방법 및 랩퍼 에이전트의 생성, 그에 따른 특성화된 랩퍼 에이전트(wrapper agent)에 대해 유효성을 검사하는 방법론 등을 제시하였다. 또한, 학습 시 필요한 지식(knowledge)으로서의 디렉토리 구성은 미리 만들어진 표준 카테고리(category)와 용어(terminology) 존재 하에 제한적이거나 새로운 디렉토리(directory) 요소에 대해 자동으로 확장할 수 있는 방법론을 제안하였다.

참고문헌

[1] J.Cowie and W. Lehnert, “ Information Extraction” Communication of the A CM, vol139, no.1, pp80-101, 1996

[2] Nicolas Kushmerick, Daniel S. Weld, and Robert Doorenbos, “ Wrapper Induction for Information Extraction” In International Joint Conference on Artificial Intelligent(IJCAI), Nagoya, Japan, 1997.

[3] R.Doorenbos, O. Etzioni, D .weld. A scalable comparison-shopping agent for the World -Wide-Web. In Proc. Autonomous Agents, 1997

[4] Bargainfinder, <http://bf.cstar.ac.com/bf/>

[5] Jango, <http://www.jango.com/>

[6] 이건명, 전자 상거래를 위한 다중 에이전트 기반 비교쇼핑 시스템 구조, 첨단정보기술연구센터, 2000.7

[7] 양재영, 최중민, 김중배, 비교 쇼핑 에이전트 시스템, HCI2000 학술대회, 2000

[8] G.Zacharia, A.Moukas, R.Guttman, P.Maes, An agent system for comparative shopping at the point of sale, MIT Meida Lab., URL://ecommerce.media.mit.edu/, 1997.

[9] Nicholas Kushmerick, Wrapper Construction for Information Extraction. PhD thesis, Univ. of Washington, 1997. In preparation