

# 전자상거래에서 비교구매 에이전트를 위한 웹 온톨로지에 대한 연구

김수경\*, 안기홍\*

\*한밭대학교 정보통신전문대학원 컴퓨터공학과  
e-mail:{kimsk, khahn}@hanbat.ac.kr

## A Study of Automatic Web Ontology for Comparison-Shopping Agent in e-Business

Su-Kyoung Kim\*, Young-Geun Kim\*, Kee-Hong Ahn\*  
Dept. of Computer Engineering, Hanbat National University

### 요 약

기존 전자상거래 플랫폼과 콘텐츠는 데이터의 확장 및 통합이 고려되지 않은 HTML을 중심으로 한 표현 기반 기술로 되어 있고, 특히 전자상거래 사이트별로 상품 정보에 대한 분류체계가 상이함으로 인해 구매자에게 상품별 비교와 검색에 있어서 많은 시간이 낭비되고 또한 정보의 공유가 어려운 기술로 인하여 판매자와 구매자 모두의 요구를 만족시키지 못하고 있다. 본 논문에서는 최근 차세대 웹 기술로 각광받고 있는 시맨틱 웹 기반 기술인 RDF/RDFS를 이용하여 기존 이종의 상점에 제시된 상품 정보를 RDF 문서로 생성하고, OWL을 사용하여 상품 지식 기반 온톨로지를 구축한 뒤, RDF 문서와의 분석과 매칭을 통하여 이종의 상점에 표현된 상품을 실시간으로 비교 검색하는 시스템을 설계 제안하였다.

### 1. 서론

기존의 HTML과 HypterText를 기반으로 한 전자상거래 상점은 구매자에게 찾고자 하는 정보의 정확한 위치와 경로 파악에 큰 어려움이 있다. 구매자가 상품을 비교 검색하는 데 있어서 한 상점에서 행한 검색 과정을 다른 상점에서도 유사하게 반복적으로 수행하여야 하며, 상점 별로 상품에 대한 정보 표현이 상이할 경우에는 검색과 비교에 더 큰 어려움이 있다. 이런 문제점을 해결하고자 비교구매 에이전트(Comparison-Shopping Agent, Shopbot)가 개발되었다. 비교 구매 에이전트는 실시간 검색 또는 제품정보 DB 형태로 개발이 되고 있으나 국내에서는 실시간 검색이 아닌 제품 데이터베이스 기반 검색으로 제공된다. 그리고 현재 대부분의 비교구매 에이전트는 가격을 포함한 상품 사양의 일부 정보만을 수집가능 할 뿐, 전체 사양정보를 완벽히 수집하고 있지 못하여 상품 간 완전한 비교가 어려울 뿐더러 지능형 구매지원을 하지 못하므로 구매자의 다양한 요구를 만족시키지 못하고 있고 구매자가 상품을

대한 정보를 정확하게 요구하지 못할 경우에는 그 효용성이 저하되는 문제가 발생되고 있다. 본 논문에서는 전자상거래 상점별로 상이하게 정의되고 표현되어 있는 디지털 캠코더에 대한 HTML상의 상품 정보를 시맨틱 웹 기반 기술인 RDF 문서로 변환한 뒤, OWL을 이용하여 디지털 캠코더(DCC: Digital Camcoder) 도메인 온톨로지 저장소(Domain Ontology Repository)에 저장된 DCC 지식 베이스 온톨로지와 비교와 추론을 통해 기존의 비교구매 에이전트의 문제점을 극복하고 동적으로 온톨로지를 재 정의하는 시스템을 제안하고자 한다.

### 2. 관련연구

#### 2.1 전자상거래와 비교 구매 에이전트

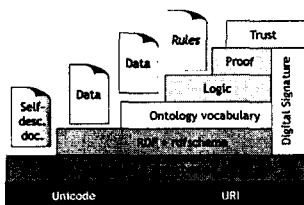
현재 대부분의 전자상거래 서비스는 구매자가 물건의 모습이나 사양을 보고 선택하는 상품 카탈로그 수준의 상태에 있다. 구매자는 자신이 원하는 물건의 사양을 간단히 표시하여 그 사양에 맞는 상품을 여러 업체로부터 제시 받기를 원하고 있으나 그것을

지원하는 시스템은 매우 적다. 그리하여 구매자가 특정 상품을 찾고자 할 때 에이전트가 구매자를 대신하여 인터넷에 존재하는 상점을 검색하여 상점별로 해당 상품을 비교하는 시스템인 비교 구매 에이전트가 앤더슨 컨설팅의 Bargain Finder에 의해 시작되었다. 그것은 미리 등록된 9개의 인터넷 상점의 상품에 대한 상품 정보 DB를 구축하여 구매자가 특정 음반 아티스트와 음반 제목을 입력하면, 여러 CD 상점의 판매 가격을 소비자에게 실시간으로 제시하는 기능이었지만 CD 판매가격이 상대적으로 비싼 상점들이 비교구매 에이전트의 정보 수집을 거부 (Agent blocking) 하면서 상업적으로 실패하게 되었다. 이러한 거부 문제를 해결하기 위해 워싱턴 대학에서 개발된 Jango라는 에이전트는 구매자의 PC에 클라이언트 소프트웨어를 설치하여 인터넷 상점의 접속을 구매자의 컴퓨터에서 할 수 있도록 하여 구매자가 많은 인터넷 상점의 가격을 비교할 수 있게 하였으나 상점의 분석 성공률이 50% 이하로 나타나는 등의 문제로 서비스가 중단이 되었다.

현재 운영되는 대부분의 국내외 비교구매 에이전트는 Bargain Finder와 같은 방법으로 제공되고 있으며, 상품 간 비교에 있어서 상품 정보가 특정 상점에는 존재하나 다른 상점에는 존재하지 않는다면 비교 자체가 불가능해지고, 또한 구매자의 다양한 요구를 만족시켜주는 지능형 구매 지원을 하지 못하는 등의 문제점을 갖고 있다.

## 2.2 시맨틱 웹과 웹 온톨로지

시맨틱 웹은 웹상의 정보에 잘 정의된 의미를 부여함으로써 사람뿐만 아니라 컴퓨터도 쉽게 문서의 의미를 해석할 수 있도록 하여 컴퓨터를 이용한 정보의 검색 및 해석, 통합 등의 업무를 자동화하기 위한 목적으로 제안되었다. 온톨로지는 특정 개념에 대한 의미를 표현하기 위해 개념과 개념들 간의 관계를 이용한 지식 표현 방식이다. 웹 온톨로지(Web Ontology)란 웹상에 존재하는 Resource들에 대한 온톨로지로서 W3C 표준인 RDF를 기반으로 정의된 온톨로지 언어를 통해 기술된다.



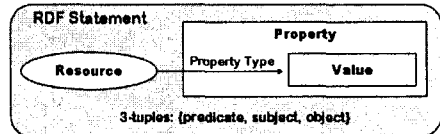
[그림1] 시맨틱 웹의 계층적 구조

[그림1]은 시맨틱 웹의 계층구조를 나타낸다. 현재는 웹 온톨로지언어로 OWL 권고안을 발표하였으며, Rule을 표현하기 위한 표준 작업을 진행하고 있다.

## 2.3 웹 온톨로지 언어(Web Ontology Language)

### (1) RDF와 RDFS

RDF(Resource Description Framework)는 W3C의 가장 기본적 시맨틱 웹 언어로서 웹에 있는 자원에 관한 메타 데이터를 표현하기 위한 언어이다. RDF 모델은 기본적으로 자원(Resource), 특성(Property), 서술문(Statement)의 개념으로 구성된다. 서술문은 일반 문자의 주어(subject), 동사(predicate), 목적어(object)에 해당하는 것으로서 사람이나 웹 문서 등 특정 대상(object)이 가지고 있는 상태를 표현하며 이것이 RDF 문의 기본 단위가 된다.



[그림2] RDF의 데이터 모델

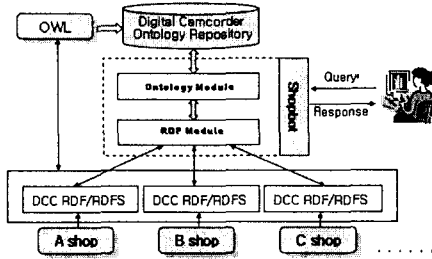
RDF의 새로운 용어를 정의하기 위해서 RDF 스키마(Schema)를 사용한다. RDF 스키마는 특성에 대한 정의나 사용상의 제약 사항을 기술한 것으로 RDF 문을 구성하는 단어(term)을 정의하고 단어들의 세부적인 의미를 기술하고 있다.

### (2) OWL

OWL은 DAML+OIL를 기반으로 발전된 형태로 시맨틱 웹을 위한 W3C의 웹 온톨로지 표준 언어이다. OWL은 DAML+ OIL과 유사한 형식을 가지며, DAML+OIL의 네임 스페이스와 속성, 클래스 이름 등을 변경하고 RDF/RDFS의 변화를 수용하였다. OWL은 DL (Description Logic)을 기반으로 만들어진 RDF 확장 언어이며, OWL의 axioms는 DAML+OIL의 axioms 보다 더욱 풍부한 표현력을 가지고 있으며, 클래스나 속성 간의 명제(subsumption)나 equivalence 등의 다양한 성격을 선언하는데 사용된다. OWL은 XML 스키마의 모든 데이터형을 지원하며 RDFS와 밀접하게 연관되어 있다. OWL은 현재 OWL Lite, OWL DL, OWL Full로 나누어져 있다.

## 3. 시스템 제안

본 시스템은 [그림3]과 같이 전자상거래 상점별로 상이하게 정의된 상품에 대한 정보를 각각의 RDF

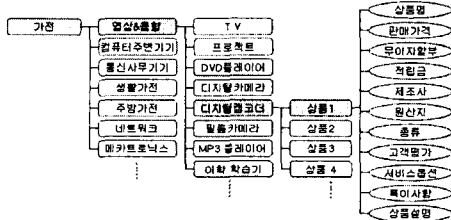


[그림3] 온톨로지 생성 시스템

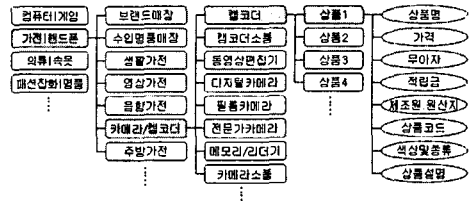
문서로 구현하였다. 그리고 OWL을 이용하여 디지털 캠코더에 대한 DCC 지식베이스 온톨로지를 정의하여 온톨로지 저장소에 저장하였다. 각 중개인의 RDF에 정의된 메타데이터를 분류한 뒤 DCC 지식베이스와의 비교 분석을 거쳐 DCC 지식베이스 온톨로지에 정의된 값(value)보다 좋은 조건일 경우 DCC 지식베이스 온톨로지의 값을 동적으로 재정의하여 DCC 도메인 온톨로지 저장소에 저장하고 구매자에게는 DCC 지식베이스 온톨로지의 조건에 가장 적합한 상점의 정보를 제시하도록 설계하였다. 이렇게 동적으로 정의된 온톨로지는 상점별로 메타데이터의 표현이 상이하게 되어 있더라도 온톨로지를 이용함으로써 동일한 의미를 가진 상품 정보의 비교와 검색에 이용될 수 있다.

3.1 각 상점별 DCC RDF

DCC를 판매하는 전자상거래 상점의 상품에 대한 분류와 정보 표현은 [그림4, 5]과 같이 상점의 상품 정보 표현에 따라 많은 차이가 있다. 또한 이를 바탕으로 설계된 RDF 문서도 웹상에 정의되어 공유되는 메타데이터 어휘를 이용하지 않을 경우 [그림6, 7]과 같이 같은 상품 정보임에도 다른 메타데이터를 갖는 문서로 작성될 수 있다. 제안된 설계는 그러한 메타데이터를 공유하지 않고 온톨로지를 이용함으로써 이러한 문제를 극복하게 된다.



[그림4] L상점의 상품에 대한 분류와 정보



[그림5] C상점의 상품에 대한 분류와 정보

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://www.dcc.com/1.0dc/" xmlns:dcc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="<dc:DigitalCamCorder>">
    <dc:title>[04년 신제품]삼성 디지털 캠코더 VM-790C</dc:title>
    <dc:publisher>LG eShop</dc:publisher>
    <dc:date>2004-8-20</dc:date>
    <dc:ProductName>삼성 디지털 캠코더 VM-D790</dc:ProductName>
    <dc:SalePrice>899,000</dc:SalePrice>
    <dc:Installation>3</dc:Installation>
    <dc:Save>0원</dc:Save>
    <dc:Maker>삼성</dc:Maker>
    <dc:Origin>한국</dc:Origin>
    <dc:Kind>-</dc:Kind>
    ...
  </rdf:Description>
</rdf:RDF>
```

[그림6] L상점의 상품(자원) RDF의 예

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://www.dcc.com/1.0dc/" xmlns:dcc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="<dc:DigitalCamCorder>">
    <dc:publisher>CJ mail</dc:publisher>
    <dc:date>2004-8-20</dc:date>
    <dc:ProductName>삼성 디지털 캠코더 VM-D790</dc:ProductName>
    <dc:Price>790,000</dc:Price>
    <dc:nonInterest>8</dc:nonInterest>
    <dc:Fund>23,700원</dc:Fund>
    <dc:MakerOrigin>삼성/한국</dc:MakerOrigin>
    <dc:ColorKind>-</dc:ColorKind>
    <dc:Option>
    ...
  </rdf:Description>
</rdf:RDF>
```

[그림7] C상점의 상품(자원) RDF의 예

[표1]은 두 상점 사이의 동일한 상품 정보이지만 서로 다른 메타데이터로 정의된 상품에 대한 메타데이터를 정리하였고 이를 기준으로 온톨로지의 클래스와 속성을 설계하였다.

[표1] 두 RDF에서 동일한 메타데이터 추출

| DCC 자원(Resource)에서 동일한 속성(Property) |                    |                    |                    |
|-------------------------------------|--------------------|--------------------|--------------------|
| 속성 값                                | L사 속성유형            | C사 속성유형            | 속성값                |
| [04년 신제품]삼성 디지털 캠코더                 | Title              | Title              | 없음                 |
| LG eShop                            | Publisher          | Publisher          | CJ mail            |
| 삼성 디지털 캠코더 VM-D790                  | ProductName        | ProductName        | 삼성 디지털 캠코더 VM-D790 |
| 899,000                             | SalePrice          | Price              | 790,000            |
| 3개월                                 | Installation       | nonInterest        | 6개월                |
| 0                                   | Save               | Fund               | 23,700원            |
| 삼성                                  | Maker              | MakerOrigin        | 삼성/한국              |
| 한국                                  | Origin             | 없음                 |                    |
| 검정/실버                               | Kind               | ColorKind          | 검정/실버              |
| 최저가 신고 대상상품                         | Option             | Option             | 배송5일 이내            |
| 상품에 대한 규격                           | ProductExplanation | ProductDescription | 상품에 대한 규격          |

3.2 상점의 DCC 온톨로지 구조 및 DCC 지식베이스 웹 온톨로지 정의

DCC 지식베이스 온톨로지 구조는 [그림8]과 같이 최상의 클래스로서 publisher을 그 아래 product 클래스와 각 속성과 값들을 설계하였다. [그림9]는 설계된 구조를 바탕으로 OWL을 이용하여 정의한 DCC 지식베이스 온톨로지의 일부이다.

