

시멘틱 웹 기반의 이미지 검색을 이용한 비교 쇼핑 시스템

이기성^o 유영훈 조근식

인하대학교 컴퓨터공학부

{funvirus^o, yhyu}@eslab.inha.ac.kr, gsjo@inha.ac.kr

Comparison Shopping System using Image Retrieval on the Semantic Web

Kee-Sung Lee^o Young-Hoon Yu Geun-Sik Jo
School of Computer Science & Engineering, Inha University

전자상거래의 발달로 인하여 여러 쇼핑물들의 상품정보를 효과적으로 비교할 수 있도록 하기위한 다양한 방법들이 연구되어져 왔다. 특히, 비교구매 쇼핑물은 사용자가 찾고자 하는 상품의 정보들을 정확히 알고 있는 상태에서 검색 조건들의 입력을 통해, 해당 상품을 보유한 쇼핑물들의 상품 정보들을 비교함으로써 보다 저렴한 상품의 구매가 이루어지도록 한다. 그러나 이러한 시스템은 원하는 상품에 대한 정확한 지식이 있는 사용자에게 유용하며, 만일 고객이 원하는 상품에 대한 정확한 지식이 없을 경우, 비교 구매 시스템의 효용성은 떨어질 수 밖에 없는 문제를 가지고 있다. 이러한 문제의 해결을 위해 본 논문은 상품에 대한 지식이 없는 사용자가 카테고리나 키워드로 검색을 하지않고, 온톨로지를 기반으로한 이미지 쿼리에 의해 결과물을 얻을 수 있도록 이미지 검색에 의한 비교 쇼핑 시스템을 제안한다. 각 쇼핑물의 상품 이미지들의 메타데이터 안에 도메인 전문가에 의해 온톨로지 기반의 daml로 생성된 주석이 추가된다. 사용자들은 이렇게 생성된 이미지들을 드래그 앤 드롭(Drag and Drop)을 통해 기존의 쇼핑물에서 복잡한 키워드로 검색 하는 것을 대체하게 되고 상품들에 대한 비교정보를 얻을 수 있다. 본 논문은 의류상품을 이용한 이미지 검색 비교 구매 시스템(Image Retrieval Comparison Shopping)을 구현하였다.

1. 서론

인터넷의 대중화와 함께 새로운 비즈니스의 환경인 전자상거래가 급속히 발전하였다. 이에 따라 시간과 공간의 제약 없이 원하는 상품을 손쉽게 구입할 수 있는 인터넷 쇼핑물이 등장 하게 되었으며, 국내 7백대 기업의 87%가 전자상거래를 이용하고 있을 정도로 그 성장 가능성은 무한하다 [1]. 그러나 하이퍼공간(Hyperspace)의 방대한 정보들의 연결은 사용자들에게 찾고자 하는 정보의 정확한 위치와 경로 파악에 큰 어려움을 준다 [2][3]. 또한 A라는 쇼핑물에서 사용자가 찾은 상품을 다른 쇼핑물들의 상품과 비교하기 위해서 사용자는 A라는 쇼핑물에서 행했던 일련의 검색 과정들을 반복적으로 수행해야 하는데, 이를 해결하기 위해 비교 구매(Comparison Shopping)시스템이 제안되었다 [4]. 하지만 기존의 비교 구매 시스템 역시 자신의 요구를 정확하게 표현할 수 없는 인터페이스와 원하는 상품에 대한 정확한 지식이 없을 경우, 질적으로 입증되지 않은 방대한 양의 정보로 인해 그 효용성이 저하되는 문제가 발생될 수 있다.

이러한 문제의 해결을 위해 본 논문은 각 쇼핑물들의 상품 이미지(jpeg) 파일속에 온톨로지(Ontology)기반의 daml로 상품의 정보 및 디자인에 대한 정보를 삽입 한다. 사용자들은 카테고리나 키워드로 검색하지 않고 이렇게 생성된 이미지들을 이용하여 여러 쇼핑물의 상품에 대한 비교정보를 얻을 수 있는 이미지 검색을 이용한 비교구매 시스템을 제안한다.

2. 관련 연구

본 논문에서 제안하는 시스템을 위한 기반기술에 관해 알아 본다.

2.1 시멘틱 웹 (Semantic Web)

시멘틱 웹이란 데이터에 의미를 부여하여 컴퓨터가 이해할 수 있는 언어로 만들어 컴퓨터에 의해 처리 될 수 있도록 고안된 차세대 웹의 비전이다 [5]. 이를 이루는 핵심적인 구성 요소로는 도메인 지식을 공유하기 위한 온톨로지(RDFs, daml+oil,OWL), 그에 따라 구조화된 데이터를 표현하기 위한 메타데이터(RDF, DAML), 자동적이고 지능적인 기능을 수행하기 위한 에이전트, 그리고 개념화되고 구조적인 데이터 기반의 추론 엔진이라고 할 수 있다.

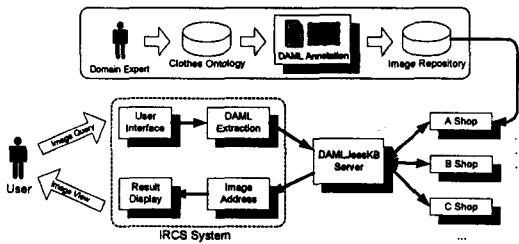
2.2 이미지 검색 (Image Retrieval)

이미지 검색 방법에는 내용 기반(Content-Based)의 이미지 검색과 키워드 기반(Keyword-based)의 이미지 검색(Image Retrieval)으로 나눌 수 있다. 내용 기반의 이미지 검색은 특징 추출(Feature Extraction)과 객체 인식(Object Recognition)의 과정으로 이루어지는 반면, 키워드 기반의 이미지 검색 방법은 우선 이미지에 대한 주석을 표현하고, 표현된 주석을 기반으로 이미지를 검색하는 방법으로써, 비교적 구현이 쉽고, 정확률이 높은 장점이 있으나, 주석을 표현할 때 개인의 주관이

포함되는 문제와 동의어(Synonyms)로 인한 재현율 (Recall) 수치와, 동음이의어(Homonyms)로 인한 정확도(Precision) 수치도 현저히 낮아지는 문제를 가지고 있다 [6]. 이러한 단점을 해결하기 위해, 본 논문에서는 시멘틱 웹 기반 기술인 daml+oil로 온톨로지를 구성하고, 단순한 키워드가 아닌 daml로 주석 처리 함으로써 검색시 이해하고 추론할 수 있게 하고자 한다.

3. 시스템 구조도 (System Architecture)

본 논문이 제시하는 시멘틱 웹 기반의 이미지 검색을 이용한 비교쇼핑 시스템 구조는 [그림 1]과 같다.



[그림 1] IRCS(Image Retrieval Comparison Shopping) 시스템 구조도

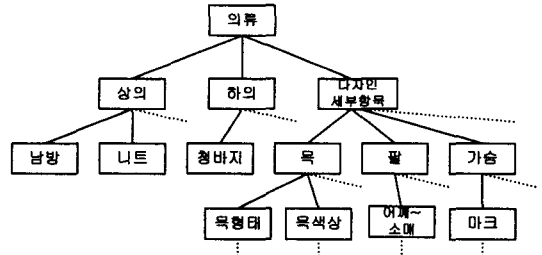
도메인 전문가는 상품 이미지들의 메타데이터 안에 온톨로지 기반의 daml형태로 어노테이션(annotation) 한다. 어노테이션의 내용은 상품의 속성값들(색상, 가격, 생산지 등)과 의류의 디자인 정보를 포함하게 된다. 사용자들이 쇼핑 중 관심 있는 상품의 이미지(어노테이션된)를 발견한다. 사용자는 이 상품을 다른 사이트에 있는 동일한 또는 유사한 상품들의 정보와 비교하고자 할 때, 본 논문에서 제안한 시스템(IRCS)의 사용자 인터페이스(User Interface)를 통해 이미지로 쿼리를 하게 된다. DAML Extraction 모듈에서 이미지 메타데이터 안에 들어있는 daml을 추출한다. 추출된 daml을 DAMLJessKB 서버로 넘겨준다. DAMLJessKB에서는 추출된 daml과 동일한 이미지 또는 유사한 이미지를 찾아서 그 결과를 daml형태로 Image Address 모듈에게 되돌려준다. Image Address 모듈에서는 daml 내용 중 이미지의 주소값을 다시 추출하게 되고 추출된 이미지의 주소값을 Result Display로 넘겨주게 된다. Result Display에서는 이미지 주소에 있는 상품의 정보들을 사용자가 보기 편한 인터페이스로 보여주게 된다.

3.1 의류 온톨로지 (Clothes Ontology)

일반적으로 온톨로지는 도메인 내의 개념 및 개념 사이의 관계, 개념의 속성, 속성에 부여된 제약조건 및 객체들로 표현된다 [7]. 이러한 온톨로지를 통해서 특정 도메인의 언어를 공통으로 정의하고, 지식을 공유할 수 있다. 본 논문에서 비교 쇼핑 시스템을 위한 도메인으로써 의류에 대한 daml+oil로 표현된 온톨로지는 [그림 2], 구조는 [그림 3]같이 구성하였다.

```
<daml_owl:Class rdf:ID="니트" />
<rdf:subClassOf rdf:resource="#상의" />
</daml_owl:Class>
<daml_owl:DatatypeProperty rdf:ID="상품명" />
<daml_owl:Domain rdf:resource="#상의" />
<daml_owl:Range rdf:resource="http://eslab.inha.ac.kr/clothesString" />
</daml_owl:DatatypeProperty>
<daml_owl:ObjectProperty rdf:ID="디자인세부사항" />
<daml_owl:Domain rdf:resource="#상의" />
<daml_owl:DatatypeProperty rdf:ID="목형태" />
<daml_owl:Domain rdf:resource="#디자인세부사항" />
</daml_owl:DatatypeProperty>
```

[그림 2] 의류 온톨로지(daml+oil)



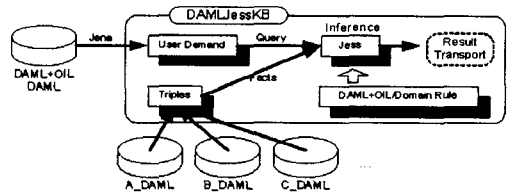
[그림 3] 의류 온톨로지 구조

3.2 이미지 주석달기 (Image Annotation)

이미지 검색을 위한 이미지 주석처리 과정은 해당 도메인에 대한 충분한 지식과 경험을 보유한 전문가들에 의해 이루어진다. 온톨로지를 기반으로 주석을 쉽게 달수 있도록 해주는 대표적인 응용프로그램은 RDFpic 이다 [8]. 이 응용프로그램은 RDF Schema를 기반으로, RDF가 Jpeg 이미지의 메타데이터로 삽입이 될 수 있도록 해주는데, 본 논문은 RDF Schema/RDF보다 의미표현이 풍부한 DAML+OIL/DAML을 사용함으로써 웹사이트들 간에 보다 높은 차원의 상호 운용성을 구축하도록 설계되었다.

3.3 질의 처리 과정 (Query Processing)

제안된 시스템의 daml Extraction 모듈을 통해서 사용자가 찾고자하는 상품의 정보(daml)를 추출하고, 각 쇼핑몰들의 상품의 정보(daml)를 DAMLJessKB [9]를 통해 JESS의 fact로 넘겨준다. 그 후 daml+oil rules/domain rules을 기반으로 추론을 수행하게 된다. 그 결과는 다시 DAMLJessKB를 통해 daml 포맷 형태로 얻게된다 [그림 4].



[그림 4] 질의 처리 과정 흐름도

4. 구현

기존의 비교구매 시스템에서 상품정보를 비교하려고 키워드를 입력할 때, 어떠한 키워드로 검색을 해야 할지 모르는 경우가 있다.

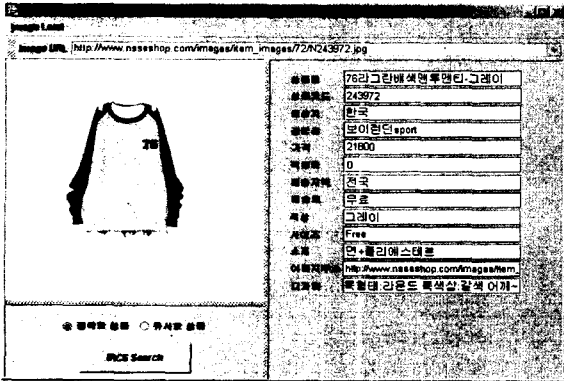
```
<clothes:니트 rdf:ID="의류온톨로지_000001" />
<clothes:상품명>762|그린 넥서 연루엔티 - 그레이|</clothes:상품명>
<clothes:상품코드>243972</clothes:상품코드>
<clothes:원산지>한국</clothes:원산지>
...
<clothes:디자인세부사항 rdf:ID="의류온톨로지_100001" />
<clothes:목>라운드</clothes:목>
<clothes:목색상>갈색</clothes:목색상>
<clothes:어깨-소매색상>갈색</clothes:어깨-소매색상>
```



[그림 5] daml로 주석 처리된 이미지 메타데이터

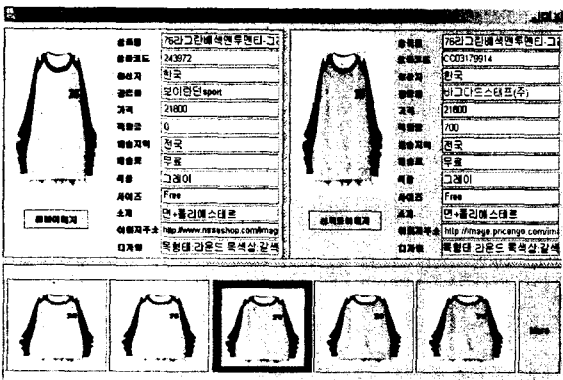
본 논문에서는 사용자 중심의 비교구매 시스템을 설계하여

사용자들이 편리하게 검색할 수 있게 한다. 만약 사용자가 [그림 5]의 상품들(이미지에 daml로 주석처리된) 다른 사이트의 상품들과 비교하려면, 사용자 인터페이스로 이미지를 드래그 앤 드롭하거나, 이미지의 주소를 입력, 또는 이미지를 불러오는 것으로 자신이 찾고자하는 상품의 정확한 정보들을 표현 할 수 있게 된다 [그림 6].



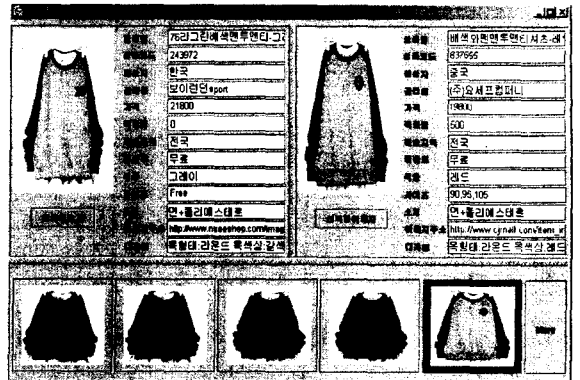
[그림 6] IRCS 시스템 사용자 인터페이스

사용자 인터페이스에서 이미지나 이미지의 주소가 들어오면 이미지의 메타데이터를 읽고, daml로 표현된 상품정보를 이용해서 사용자가 볼 수 있도록 해준다. 사용자는 동일한 디자인의 제품을 검색할 것인지, 유사한 디자인의 제품을 검색할지 선택한 후 검색을 시작한다. 동일한 디자인 검색시 디자인 정보의 값과 다른 사이트에 있는 제품의 정보를 DAMLJessKB에서 daml+oil rules/domain rules을 가지고 Jess추론 엔진을 통해서 디자인 속성값이 동일한 제품만을 한 화면에 5개씩 Result Display에서 보여 주게 된다 [그림 7].



[그림 7] IRCS 시스템 결과 화면 (동일한 디자인)

그리고 검색된 상품 이미지를 선택하면 원본 상품정보와 선택된 상품정보의 내용이 같은 것은 검정색, 다른 것은 파란색으로 결과가 나타남을 볼 수 있다. 유사한 디자인 검색시 디자인 정보의 값과 다른 사이트에 있는 제품의 정보를 DAMLJessKB에서 daml+oil rules/domain rules을 가지고 Jess 추론 엔진을 통해서 가장 유사한 순서대로 5개씩 Result Display에서 보여주게 된다 [그림 8].



[그림 8] IRCS 시스템 결과 화면 (유사한 디자인)

5. 결론 및 향후 연구

기존 비교구매 시스템은 원하는 상품에 대한 정확한 지식이 있는 사용자에게 유용하며, 만일 고객이 원하는 상품에 대한 정확한 지식이 없을 경우, 비교 구매 시스템의 효용성은 떨어질 수 밖에 없는 문제를 가지고 있다. 그러나 제안한 시스템을 통해서 사용자들은 상품정보에 대해 정확한 정보를 몰라도 이미지(은플로지 기반의 daml로 주석이 삽입 되어있는)만 가지고 찾고자하는 상품 정보를 정확하게 찾을 수 있었다. 향후 연구로는, 이미지 주석처리를 위해 소비되는 시간을 줄이기 위한 방법과 이미지의 메타데이터에서 daml을 추출하고 비교하는 시간을 효과적으로 줄이기 위한 방법에 대한 연구를 진행 할 것이다.

감사의 글

본 연구는 한국과학재단 지정 인천대학교 동북아 전자 물류 연구센터의 지원에 의한 것이다.

6. 참고 문헌

- [1] <http://www.ibm.com/kr>, 한국IBM, 2003
- [2] IDC 보고서(1996). 전자상거래 빠르게 증가. 컴퓨터 월드, 8월15일자.
- [3] Woodhead, Nigel. "Hypertext and Hypermedia : Theory and Applications", London : Sigma Press and Addison-Wesley Co. 231 p.
- [4] R.Doorenbos, O. Etzioni, and D. Weld. "A Scalable Comparison-Shopping Agent for the World-Wide Web." In Proceeding of the First International Conference on Autonomous Agent, 1997.
- [5] Tim Berners-Lee, James Hendler and Ora Lasilla. "The Semantic Web", The Scientific American, May 2001
- [6] Eero Hyvönen, Samppa Saarela, Avril Styrman, Kim Viljanen. "Ontology-Based Image Retrieval" IW3C2, 2003.
- [7] Noy, Natalye Fridman, and Deborah McGuinness. "Ontology Development 101: A Guide to Creating Your First Ontology," Knowledge Systems Laboratory, Stanford University 2001.
- [8] rdfpic. <http://www.w3.org/TR/photo-rdf/>, 2002
- [9] <http://edge.cs.drexel.edu/assemblies/software/damljesskb/damljesskb.html>