

온톨로지 매칭을 이용한 RSS 기반의 비교쇼핑 시스템*

박 상 언**

<목 차>

I. 서론	4.1 RSS 기반 비교쇼핑 시스템 구조와 프로세스
II. 관련연구	4.2 비교쇼핑을 지원하기 위한 RSS 피드의 내용
2.1 비교쇼핑 관련연구	4.3 온톨로지 매칭 알고리즘
2.2 Web 2.0과 RSS의 전자상거래 활용	V. RSS 기반 비교쇼핑 시스템의 상품매칭 성능 실험
2.3 온톨로지 매칭	VI. 결론
III. RSS를 이용한 비교쇼핑 방안 및 대안 분석	참고문헌
3.1 비교쇼핑의 한계와 RSS 활용의 장점	<Abstract>
3.2 비교쇼핑에서의 RSS 활용 이슈와 대안	
IV. 온톨로지를 이용한 RSS 기반 비교쇼핑 시스템	

I. 서론

인터넷으로 상품을 구매하기 위해서는 구매자가 여러 판매자들이 제시하는 상품들의 판매 조건과 속성에 대한 정보를 수집하고 이 정보들을 비교 평가함으로써 최선의 상품을 선택하게 되는데, 여기에는 많은 노력과 시간이 요구된다. 비교쇼핑 서비스를 제공하는 가격비교 사이트를 이용하면 가격 등의 결제조건을 포함한 상품 속성들과 상품을 제공하는 판매자들을 보다 쉽게 비교할 수 있으나, 원하는 물품이 특정 쇼핑몰에

있음에도 불구하고 검색이 되지 않거나 상품 정보가 부정확한 경우가 빈번히 발생하고 있다(권혁주, 2008). 또한 지나치게 많은 결과가 나오기 때문에 여전히 비교에 어려움을 겪는 경우도 많다.

본 논문에서는 RSS를 이용하여 이러한 비교쇼핑의 단점들을 극복하기 위한 방안을 제시하고자 한다. RSS는 웹 2.0의 대표적인 기술 중 하나로, 뉴스나 블로그와 같이 자주 업데이트되는 콘텐츠를 직접 방문하지 않고도 자신의 컴퓨터 혹은 웹 사이트에서 구독할 수 있게 하는 표준

* 본 연구는 2009학년도 경기대학교 학술연구비(일반연구과제) 지원에 의하여 수행되었음.

** 경기대학교 경영정보학과 조교수

이다. RSS는 현재 아마존(Amazon.com), 바이닷컴(Buy.com) 등의 외국 쇼핑 사이트 외에 G마켓(www.gmarket.co.kr), 11번가(www.11st.co.kr), 예스24(www.yes24.com), 알라딘(www.aladin.co.kr) 등의 국내 쇼핑 사이트에서도 활발하게 사용되고 있는데, 쇼핑물에서 상품정보에 대해 RSS를 제공하면, 사용자는 각 사이트에 방문하지 않고도 자신이 원하는 상품들에 대한 정보를 받아보는 것이 가능하다. RSS는 각 쇼핑물에서 직접 제공되기 때문에 비교쇼핑 사이트와는 달리 항상 정확한 정보를 담고 있어 완전성과 부정확성의 문제를 해결할 수 있고, 사용자가 관심 있는 쇼핑물에서 원하는 상품의 범위를 지정하여 RSS 피드를 만들기 때문에 사용자가 관심 없는 상품들의 수를 줄이고 개인화된 결과를 제공할 수 있는 것이 가능하다.

그러나, 서로 다른 여러 쇼핑물에서 제공하는 RSS 피드의 상품들 중에서 같은 분류에 있는 상품들끼리만 모아 비교하기 위해서는 서로 다른 형식과 내용으로 작성된 RSS 피드의 상품 분류 및 속성을 RSS 리더기가 이해할 수 있어야 한다. 이를 위해서는 형식적인 상호운용성(syntactic interoperability)과 의미적 상호운용성(semantic interoperability)이 보장되어야 하는데, 이를 위해 본 논문에서는 각각 RSS 표준을 이용한 상품의 표현방안과 온톨로지 매칭방안을 제시하였다.

본 연구에서는 온톨로지 스키마만을 사용하는 스키마 기반 온톨로지 매칭 중에서 상품 분류의 경로를 활용하는 구조 수준(structure level)의 온톨로지 매칭을 사용하였으며, 상품분류에 대하여 워드넷(Wordnet) (Miller, 1995)에서 정확한 의미(sense)를 찾아내기 위해 UMND1

(Unsupervised Word Sense Disambiguation Using Contextual Semantic Relatedness) (Patwardhan et al., 2007)을 이용하여 온톨로지 매칭의 사전절차로 이용하였다. 그리고 실험을 통해 이러한 온톨로지 매칭 기반의 비교쇼핑의 효과를 보였다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문과 관련된 다양한 기존연구를 서술하였으며, 3장에서는 비교쇼핑에서의 RSS를 활용하는데 발생하는 이슈를 분석하고 이에 대한 대안을 정리하였다. 4장에서는 RSS 기반 비교쇼핑 구조와 프로세스를 보이고, 비교쇼핑을 지원하기 위한 RSS 피드의 내용에 대해 정리하였으며, 온톨로지 매칭 알고리즘을 제시하였다. 5장에서는 실험을 통해 본 논문에서 제시하는 RSS 기반 비교쇼핑 시스템의 상품매칭 성능을 보였다.

II. 관련연구

2.1 비교쇼핑 관련연구

인터넷으로 상품을 구매하기 위해서는 여러 판매자들이 제시하는 상품들의 판매조건과 속성에 대한 정보를 수집하여야 하며, 인터넷 쇼핑물이 많은 인기를 얻게 된 배경에는 이러한 활동의 편리성에 있다고 볼 수 있다(김성연, 2008). 이러한 정보들을 비교 평가함으로써 사용자는 최선의 상품을 선택하게 되는데, 일반적으로 사용자가 이러한 정보를 수집하기 위해 각 판매자들의 사이트를 방문하여 원하는 상품을 검색하고, 그 결과들을 정리하여 비교하는 데에는 많은 노력과 시간이 요구된다. 이러한 노력과 시간의 소모

를 줄이기 위해 제공되는 서비스가 바로 비교쇼핑이다(이한진 외, 2005).

가격비교 사이트를 이용하면 가격 등의 결제 조건을 포함한 상품 속성들과 상품을 제공하는 판매자들을 보다 쉽게 비교할 수 있다(노상규 외, 2007). 이러한 상황에서 소비자들이 가격이나 서비스 품질만 고려하지 않고 가격대비 서비스 품질(quality price ratio)을 고려해서 판매자를 선택하는 경향이 있다는 사실(Staat et al., 2002)에 근거하여 화폐가치를 기반으로 한 상대적인 평가를 반영하고자 한 연구들이 있다. 이 연구들은 상품의 비교보다는 결제서비스, 배송서비스, 이벤트 유무, 적립금 형태 등의 판매자 정보를 이용하여 최상의 판매자를 추천하고 있다(노상규 외, 2007; Talluri et al., 2006; Zhu, 2004). 그러나 이러한 연구들의 문제점은 다양한 판매자 서비스들을 평가해서 수치화하고 여기에 가중치를 부여하여 미리 만들어진 모형에 따라 각 판매자들을 평가해야 한다는 점이다. 오픈마켓과 같이 수많은 판매자가 존재하는 현재의 상황에서는 현실적으로 적용하기 어려운 방법이라고 할 수 있을 것이다.

시맨틱 웹은 개인화, 지능화, 융합 서비스를 가능하게 하며 특히 온톨로지는 시맨틱 웹의 기본 요소이다(주재훈, 2009). 이러한 시맨틱 웹 기술을 비교쇼핑에 이용한 연구로 시맨틱 웹 기반의 이미지 검색을 이용하여 키워드 검색보다 나은 비교쇼핑 검색결과를 제공하고자 한 연구가 있다(이기성 외, 2005). 이 연구에서는 상품 이미지에 온톨로지를 기반으로 하여 상품정보와 디자인 정보를 포함하는 주석을 이미지의 메타데이터에 추가하고 사용자가 관심 있는 상품의 이미지와 유사한 다른 상품들을 검색하여 추천하

였으나, 이미지 자체보다 온톨로지 기반으로 표현된 제품 속성에 의한 비교쇼핑이란 표현이 더 적절할 것으로 생각되며, 각 속성에 대하여 유사도를 계산하는 수식과 가중치를 설정하여야 한다는 점에서 역시 한계가 있다.

비교쇼핑과 관련한 비교적 최근의 연구로 비교쇼핑에 상품추천과 판촉활동을 결합하고자 한 연구가 있다(Garfinkel et al., 2008). 일반적인 비교쇼핑 에이전트들이 가격 위주의 단순한 상품비교만을 제시하는 것에 착안하여, 각 쇼핑물 사이트가 제공하는 판촉내용과 추천시스템을 비교쇼핑에 결합함으로써 구매자에게 보다 나은 서비스를 제공할 수 있으며, 개별적인 쇼핑물의 판촉활동과 구매자의 요구가 어긋난 경우에 비교쇼핑 에이전트가 중재역할을 할 수 있다고 주장하였으나, 수식을 이용한 이론적인 연구에 그치고 있어 실제로 이를 어떻게 구현할 것인가에 대한 내용이 없다는 한계가 있다. 이와 유사하게, 비교쇼핑 사이트가 다양한 가격할인 효과를 반영하지 않고 있어 비교쇼핑에서 제시한 가격이 실제 쇼핑물에서 제시하는 가격과 다른 점을 해결하고자 한 연구가 있다(Chang et al., 2010). 이 연구에서는 이러한 쇼핑물의 다양한 가격할인 정책과 사용자의 요구사항을 반영하여 가격리스트를 제공하는 모델을 제안하였으나, 앞선 연구와 마찬가지로 이를 어떻게 구현할 것인가는 또다른 연구라 할 수 있다.

이상 비교쇼핑과 관련한 기존연구와 비교하여 볼 때, 본 연구는 웹 2.0 기술과 시맨틱 웹 기술을 이용하여 현재의 비교쇼핑 사이트들이 안고 있는 문제점들을 해결하고자 하는데 그 의의가 있다고 할 수 있다.

2.2 Web 2.0과 RSS의 전자상거래 활용

웹 2.0은 기존 웹의 전환점인 새시대의 등장을 나타내기 위해 만들어진 후(O'Reilly, 2005), 웹 2.0 컨퍼런스를 통해 다양한 기술과 응용 및 사례들이 소개되면서 새로운 기술과 응용을 다루는 차세대 웹 동향으로 인식되기 시작했다. 웹 2.0의 대표적인 기술 중 RSS는 콘텐츠의 유통방식을 변화시키는 핵심요소로서, "RDF Site Summary", "Really Simple Syndication", "Rich Site Summary" 등 여러 이름으로 해석되고 있으며, 뉴스나 블로그 혹은 쇼핑물의 상품 리스트 등의 다양한 내용들을 구독할 수 있도록 만들어진 표준이다. RSS 리더기에 사용자가 원하는 RSS 주소를 입력하면, 콘텐츠에 대한 메타데이터와 링크, 간략한 소개, 혹은 콘텐츠의 전체 내용이 포함된 RSS 피드를 리더기로 구독하여 볼 수 있다.

RSS를 마케팅에 활용할 경우에는, 첫째 XML을 기반으로 개인의 기호에 맞게 내용을 생성함으로써 콘텐츠의 개인화가 가능하고, 둘째 지속적인 RSS 피드의 갱신을 통해 고객이 항상 최신의 정보를 볼 수 있게 하며, 셋째 고객의 요구에 의해 RSS 피드가 생성되었기 때문에 이메일에 비해 거부감이 덜하다는 장점이 있다(이머징테크, 2006). 이와 같이 사용자, 제공자, 마케팅 입장에서 장점은 전자상거래 환경에서의 RSS 활용에 많은 동기를 부여하고 있다.

RSS는 여러 분야에서 이미 다양하게 사용되고 있는데, 그 중에서도 뉴스(Blansit, 2006), 블로그(Karger and Quan, 2005), 멀티미디어 분야의 포드캐스팅(Rethlefsen and segovis, 2006)에서 활발하게 이용되고 있으며, 비디오의 보급

(Curran and McKinney, 2006), 의학(Cheung et al., 2005), 과학기술정보의 배급(권이남 외, 2005), 학회정보의 개인화(권이남 외, 2004)에 대한 연구가 수행된 바 있다. RSS는 쇼핑몰에서도 활발하게 사용되고 있다. 대표적인 전자상거래 사이트인 아마존(Amazon.com)은 분야별로 가장 잘 팔리는 상품에 대한 정보를 사용자에게 제공하기 위해 RSS를 사용하고 있으며, 바이닷컴(Buy.com)은 별도의 웹 페이지에 RSS 피드에 대한 안내를 모아 놓고, 사용자가 중분류 수준의 상품분류를 선택하면 해당 분류에 대해 특가상품 혹은 특별상품을 안내하는 형태로 운영하고 있다. 경매로 유명한 이베이(ebay.com)에서는 모든 상품분류에서 RSS를 지원하고 있으며, 현재 경매가 진행 중인 상품에 대한 정보를 RSS 피드로 받아볼 수 있다. 한영춘, 서영석(2009)은 웹 2.0을 기반으로 하는 인터넷 쇼핑몰을 제안하면서, RSS를 웹 2.0의 주요 기술로 소개하고 RSS 공개여부를 웹 2.0 기반 쇼핑몰의 설계에 포함하였다. 또한 RSS를 이용할 경우 판매자가 제공하는 상품 정보를 다른 사람과 공유하는데 장점이 있으므로 다고 설명하였다.

국내에서는 G마켓(www.gmarket.co.kr), 11번가(www.11st.co.kr)와 같은 오픈마켓과 예스24(www.yes24.com), 알라딘(www.aladin.co.kr) 등의 온라인 서점 등에서 RSS를 지원하고 있다. G마켓은 별도의 RSS 지원화면에서 분류와 속성을 지정함으로써 원하는 상품에 대한 RSS 피드를 구독할 수 있으며, 11번가에서는 상품분류를 통해 상품을 보는 도중에 각 분류가 있는 화면에서 RSS 피드를 등록하는 것이 가능하다. 예스24와 알라딘은 아마존이나 바이닷컴과 유사하게 RSS서비스 혹은 RSS센터 등의 전용 웹페이지

에서 분야별 베스트셀러, 분야별 신간 혹은 이벤트 등 다양한 방법으로 RSS 피드를 구독하는 것이 가능하다. 이와 같이 여러 사이트가 다양한 방법으로 RSS 서비스를 제공하고 있으나, 각기 다른 방법을 사용하고 있어 RSS 피드의 범위나 내용에 통일성이 부족하다는 단점이 있다.

이 외에도 RSS를 일반적인 전자상거래 환경에 사용하는 데에는 아직 해결해야 할 문제점들이 있는데, 일반사용자들이 RSS를 지원하는 사이트를 찾기 어렵고, 사이트마다 RSS 피드 생성 방식이 달라 혼란스러우며, RSS 피드 생성시 사용자 요구의 표현이 제한적이고, RSS 피드에서 생성된 정보들을 서로 비교하기가 어렵다는 점이 그 예이다. 이와 같은 단점을 극복하고 RSS를 일반적인 전자상거래 환경에 적용하고자 한 이전 연구(박상언 외, 2007)에서는 웹 서비스와 유사한 체계적인 형태로 RSS를 전자상거래 환경에 사용하기 위한 프레임워크를 제안하였다. 이 연구의 프레임워크를 따를 경우 사용자는 개별적인 쇼핑몰 등의 전자상거래 사이트를 다니면 일일이 RSS 피드를 등록하는 번거로움 없이 지능형 서비스를 이용해 한번에 원하는 상품에 대한 RSS 피드들을 등록하고 해당 상품들에 대한 비교를 할 수 있다는 장점이 있다. 그러나, 이를 실현하기 위해서는 광범위한 표준화 체계가 만들어져야 하며, 모든 사이트가 이 표준을 따라야 한다는 제약조건이 있다.

본 논문에서는 업계에서의 광범위한 표준을 만들어 사용하지 않는 범위 내에서 최소한의 표준으로, RSS 리더기 수준의 개인화된 비교쇼핑 결과를 제공하는 시스템을 제시하고자 한다.

2.3 온톨로지 매칭

RSS 리더기가 서로 다른 여러 쇼핑몰에서 제공하는 상품 분류 및 속성을 이해하도록 하기 위해서는 두 가지를 고려하여야 하는데, 첫째가 형식적인 상호운용성(Syntactic interoperability)이다(Veltman, 2001). 상품물에서 보내온 상품 정보들은 단순한 문자열이 아니라 구조화된 형태로 표현되어야 하며, RSS 표준과 XML은 바로 이와 같은 필요성을 만족시키기 위해 만들어졌다고 할 수 있다. 둘째 고려사항은 의미적 상호운용성(Semantic interoperability)으로 이는 사전에 합의되고 공유된 단어와 표현을 이용하여 정보를 교환하는 것을 의미한다(Veltman, 2001). 온톨로지 매칭은 의미적 상호호환성을 구현하기 위한 방법으로, 서로 다른 온톨로지를 구성하는 요소들 사이의 관계를 찾아주는 알고리즘 혹은 프로그램이다(Noy and Musen, 2003). 본 논문에서는 RSS 리더기에서 서로 다른 온톨로지를 사용하고 있는 쇼핑몰들로부터 상품 정보를 받았을 때, 온톨로지 매칭을 통해 같은 분류에 속하는 상품들을 매칭하고 있다.

온톨로지 매칭을 위한 다양한 방법론들이 연구되어 왔는데(Veltman, 2001; Benetti et al., 2002; Noy and Musen, 2003; Ehrig and Staab, 2004), 크게 공유 온톨로지를 구축하여 이에 따라 효율적인 매칭을 하는 방안과, 공유 온톨로지를 사용할 수 없는 경우 경험 혹은 학습을 통해 매칭을 하는 방안으로 나눌 수 있다. 온톨로지 매칭은 다시 매칭의 대상을 클래스에 한정하는지 아니면 인스턴스까지 고려하는지에 따라 나뉘는데, 인스턴스까지 고려하면 정확성을 높일 수 있지만 매칭 시간이 기하급수적으로 증가할

수 있다는 단점이 있다. 본 연구에서는 온톨로지 스키마만을 사용하는 스키마 기반 매칭을 이용하여 온톨로지 매칭을 구현하고자 한다(Shvaiko and Euzenat, 2005). 스키마 기반 매칭 내에서는 다시 온톨로지 매칭의 방법에 따라 두 가지로 나뉘는데, 첫째인 요소 수준(Element level)의 매칭에서는 클래스의 이름만으로 매칭을 하는 반면 둘째인 구조 수준(Structure level)의 매칭에서는 클래스 간의 관계를 분석하여 매칭을 수행한다. 상품 분류는 계층적 구조를 가지기 때문에 이를 온톨로지 매칭에 활용하는 것이 가능하다. 따라서 본 논문에서는 상품 분류의 경로를 활용하는, 구조 수준의 온톨로지 매칭을 사용한다. 특히, 본 논문은 동의어 및 유사어로 기술된 상품 분류에 대비하여 워드넷(WoReNet)(Miller, 1995)을 사용하여 의미를 확장하는 방법론을 사용하는데(Guarino, 1999) 이는 온톨로지 매칭 방법론에서 구조 저장소(repository of structures) 방식과 유사하다고 할 수 있다. 또한 UMND1 (Unsupervised Word Sense Disambiguation Using Contextual Semantic Relatedness) (Patwardhan et al., 2007)을 이용하여 워드넷에서 정확한 의미(sense)를 찾아내는 절차를 온톨로지 매칭의 사전절차로 사용하였다.

최근에는 하나의 유사도를 사용하던 이전의 연구들에 비해 온톨로지 매칭에 여러 유사도를 조합하여 동시에 사용함으로써 매칭 효과를 높이고자 노력하는 연구들이 많아지고 있다 (Alasoud et al., 2009).

III. RSS를 이용한 비교쇼핑 방안 및 대안 분석

3.1 비교쇼핑의 한계와 RSS 활용의 장점

소비자는 습관적으로 구매하는 편의품(convenience goods)이 아닌 일반적인 선매품(shopping goods)을 구매할 때, 한 판매자 내에서 여러 상품을 비교하기도 하고 동일한 상품에 대해 다양한 판매자를 비교하기도 한다 (Porter, 1974). 온라인 쇼핑이 활성화되면서 예전에 비해 구매에 필요한 상품과 판매자에 대한 정보를 얻는 것이 수월해졌지만, 온라인 상의 판매자와 상품이 늘어남에 따라 정보를 습득해야 할 대상은 오히려 기하급수적으로 늘어나고 있다. 특히 G마켓(www.gmarket.co.kr)이나 11st(www.11st.co.kr)과 같이 다수의 상점이 입점하는 오픈마켓이 늘어남에 따라 검색의 어려움은 더욱 커지고 있다.

비교쇼핑은 이러한 어려움을 일부 해결하고 있으나, 몇가지 문제점을 안고 있다. 첫째, 상품 검색의 완전성이 떨어진다. 예를 들어, 원하는 물건이 실제 G마켓에는 있으나 비교쇼핑에서는 검색되지 않는 상황이 자주 발생한다. 둘째, 상품 정보가 부정확한 경우가 발생한다. 실제로 가격 비교사이트의 가격정보 정확도를 조사한 결과, 10위권 밖의 사이트들은 정확도가 90%에도 미치지 못한 것으로 나왔으며 12위인 쇼핑몰은 72.1%의 정확도를 보였다(권혁주, 2008). 가격 외에 다른 조건들도 이러한 부정확성의 문제를 안고 있다. 셋째, 지나치게 많은 결과를 보여주고 있다. 예를 들어, 비교쇼핑 사이트에서 LCD TV를 검색하게 되면 600여개의 모델과 각 모델에 대해 많게는 100개가 넘는 상점들을 보여주고 있다.

RSS를 이용하면 이러한 비교쇼핑의 단점들을

극복하는 것이 가능하다. 우선 RSS는 각 쇼핑몰에서 제공되는 것으로 쇼핑몰에서의 정보를 그대로 가져오기 때문에, 위에서 설명한 완전성과 부정확성의 문제를 해결하는 것이 가능하다. 또한 사용자가 관심 있는 쇼핑몰에서 원하는 상품의 범위를 지정하여 RSS 피드를 만들기 때문에, 대안이 될 수 있는 상품의 수를 줄이고, 각 사용자에게 개인화된 결과를 제공한다는 점에서 장점이 있다. 쇼핑몰 입장에서도 RSS를 이용하여 사용자에게 쿠폰 등을 제공함으로써 주의와 흥미를 유발시키고 사이트로의 직접연결과 구매를 유도한다는 측면에서 매력이 있다(채혁기 외, 2008).

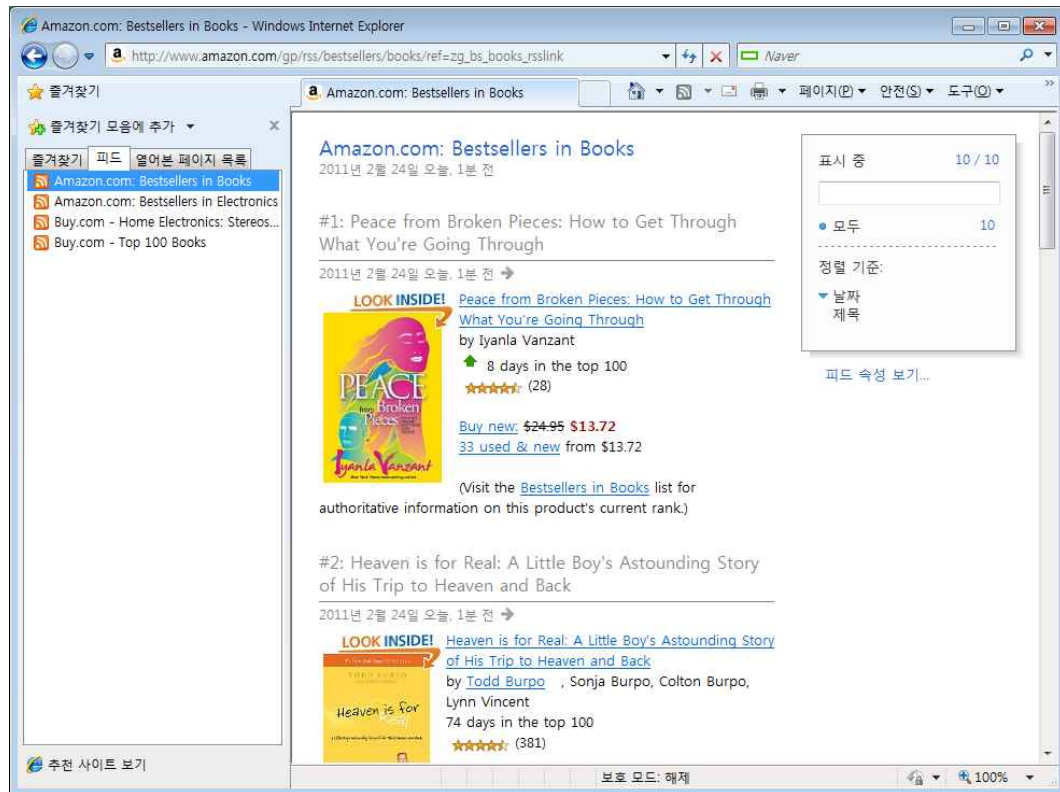
예를 들어, 사용자가 5.1 채널 스피커를 이용해 홈씨어터를 꾸미고 싶은 경우에 기존의 검색 방식에서 “5.1 채널 스피커”를 입력하면 홈씨어터를 위한 스피커 외에도 하이파이 스피커, 차량용 스피커, PC용 스피커까지 모두 검색될 뿐만 아니라 사용자가 선호하지 않는 모든 쇼핑몰의 제품들이 포함되어 나오게 된다. RSS를 이용하면, 먼저 사용자가 원하는 쇼핑몰에서 판매하는 제품 정보를 볼 수 있으며, 각 쇼핑몰에서 입력한 RSS 피드에 따라 미리 입력한 조건에 해당하는 제품들만 보는 것이 가능하다.

<그림 1>은 인터넷 익스플로러 8에서 제공하는 RSS 지원기능을 이용하여 Amazon.com과 Buy.com으로부터 관심상품에 대한 RSS 피드를 등록하고 상품 리스트를 보는 화면이다. 화면 왼쪽에 등록된 RSS 피드를 선택하면 지속적으로 업데이트되는 상품들의 리스트를 볼 수 있다. 화면 오른쪽의 정렬 기준을 이용하면 날짜와 제목 중에서 선택하여 정렬된 리스트를 볼 수 있다. 이 때 정렬 기준은 RSS 피드를 제공하는 쇼핑몰

의 정보에 따라 달라진다. 예를 들어 Amazon.com은 화면에서 보듯이 날짜와 제목에 따른 정렬을 지원하지만 Buy.com은 price와 rank에 따른 정렬을 지원한다.

쇼핑몰에서 제공하는 이와 같은 RSS 피드는 자체로 유용한 기능을 제공하지만, 비교쇼핑의 지원 측면에서는 아직 부족한 점이 많다. 첫째 서로 다른 쇼핑몰의 피드를 합쳐서 볼 수 없다. 여러 쇼핑몰로부터 RSS 피드를 받게 되면 화면 좌측에서 그 중 하나만을 선택하여 보는 것이 가능하다. 둘째 현재 쇼핑몰에서 제공하는 RSS 피드의 상품분류가 세밀하지 않기 때문에 하나의 피드에 너무 많은 상품이 나열된다. 실제 쇼핑몰에서는 보다 세부적인 상품분류를 사용하면서도 RSS 피드에서는 편의상 대분류 정도로만 분류를 제공하고 있다. 셋째 각 쇼핑몰의 정렬기준이나 제품정보가 제각각이기 때문에 서로 다른 쇼핑몰의 제품들을 비교하는 것이 불가능하다. 즉, 각기 다른 쇼핑몰로부터 온 RSS 피드에 있는 상품 정보의 형식과 내용이 상이하여 어떤 상품들을 서로 비교하여야 하는 지 알 수 없다는 것이다.

구매자가 원하는 상품분류를 소분류 수준까지 지정하고 온톨로지 매칭을 이용해 다른 판매자의 RSS 피드에 있는 유사한 상품들을 서로 매칭하면, 위에서 제기한 문제들을 해결하는 것이 가능하다. 앞서 설명한 예로 보았을 때, 사용자는 첫째 관심 있는 A사의 스피커를 각기 다른 쇼핑몰에서 어떤 조건으로 팔고 있는 지 하나의 화면에서 비교할 수 있으며, 둘째 5.1채널 안에서 더 세부적인 분류를 줄 수 있는 경우 홈씨어터를 위한 스피커에 한정하여 정보를 받아볼 수 있다. 셋째 온톨로지 매칭에서 상품분류의 계층구조를



<그림 1> 인터넷 익스플로러에서 RSS를 이용해 관심상품 보기

이용하면, 같은 5.1 채널 스피커라도 차량용과 가정용을 구분하여 비교하는 것이 가능하다. 차량용 스피커는 일반적으로 자동차용품 분류의 하부에 위치하며, 가정용 스피커는 영상음향 분류의 하부에 위치하기 때문에 두 분류의 스피커가 하나의 피드 안에 섞여 있더라도 온톨로지 매칭에서 이를 구분하는 것이 가능하다.

3.2 비교쇼핑에서의 RSS 활용 이슈와 대안

RSS 환경을 이용한 비교쇼핑을 제안하기 위하여 우선 결정하여야 할 것은 비교쇼핑을 지원하는 RSS 리더기의 역할을 어느 범위까지로 할 것인가이다. 먼저 생각할 수 있는 대안은 RSS

리더기가 사용자가 원하는 상품의 분류와 속성을 입력 받아, 온톨로지 매칭을 이용하여 먼저 각 쇼핑물에 적합한 쿼리를 생성하고, 이 쿼리를 이용해 각 쇼핑물에 RSS 피드를 요청하여 받아오는 형태이다. 그러나 이를 지원하기 위해서는 RSS 서비스의 등록 및 검색 표준화와 RSS 피드의 생성 요청 및 응답에 대한 표준화가 요구되는데 이는 현실적으로 어렵다고 할 수 있다. 따라서 본 연구에서는 사용자가 관심을 갖고 있는 상품에 대한 RSS 주소가 사용자에게 의해 RSS 리더기에 입력되어 있는 상황에서, 주어진 RSS 피드 정보를 분석하여 상품들을 비교하는 것으로 범위를 한정하고자 한다.

표준화되지 않은 상품분류 및 속성으로 각 쇼

핑몰이 보내오는 RSS 피드를 이용하여 비교쇼핑을 지원하기 위해 핵심적으로 지원되어야 할 부분은 바로 상호운용성이다. 상호운용성은 관련연구에서 살펴본 바와 같이 형식적 상호운용성과 의미적 상호운용성으로 나누어지는데, 형식적 상호운용성은 상품에 대한 정보를 담은 RSS 형식 표준으로 지원이 가능하고 의미적 상호운용성은 온톨로지 매칭으로 지원이 가능하다. 그러나, RSS 환경에서 비교쇼핑을 구현하기 위해서는 두 가지 상호운용성 모두 다음과 같이 해결되어야 할 이슈들을 갖고 있다.

첫째, 형식적 상호운용성과 관련하여 RSS 피드의 형식과 내용을 어떻게 정의하여야 할 것인지의 이슈가 있다. 본 논문에서는 주어진 RSS 표준 내에서 온톨로지 매칭을 이용해 상품 비교를 하기 위한 정보들을 표현하고자 한다. 그러나, 현재 쇼핑몰들이 지원하고 있는 RSS의 내용만으로는 온톨로지 매칭에 필요한 정보를 얻기 어렵기 때문에, 이미 제안된 적이 있는 RSS 활용방안(채혁기 외, 2008)을 이용하여 RSS 피드 내에 필요한 상품정보들을 표현하고자 한다.

둘째, 의미적 상호운용성을 위해서는 온톨로지 매칭을 이용해야 하는데, 이와 관련하여 우선 자체 온톨로지를 가질 것인지에 대한 이슈가 있다. 이는 비교쇼핑을 지원하는 방식에 따라 결과가 달라질 수 있다. 만일 사용자가 원하는 상품의 요구사항을 입력하고 이에 맞는 상품들만을 RSS 피드들로부터 가져와 비교하고 싶은 경우에는 기준이 되는 온톨로지가 필요하다. 이 때 온톨로지 매칭은 기준이 되는 온톨로지로부터 각 쇼핑몰의 상품 온톨로지로 일어나게 된다. 다른 방법으로, 사용자가 RSS 리더기에서 특정 판매자의 상품을 선택하고 이와 유사한 다른 상품

들을 검색하는 경우에는 자체 온톨로지가 필요 없을 수 있다. 이 때는 선택한 판매자의 온톨로지로부터 다른 판매자들의 온톨로지로 온톨로지 매칭이 일어난다. 본 연구에서는 두 가지 방식을 모두 지원하기 위하여 자체 온톨로지를 구축하기로 하였다.

셋째, 온톨로지 매칭을 위하여 외부 쇼핑몰의 온톨로지를 내부에 유지할 것인지의 이슈가 있다. 일반적인 온톨로지 매칭을 위해서는 온톨로지 전체가 요구된다. 그러나 이를 위하여 각 쇼핑몰의 온톨로지를 리더기 내에 유지하게 되는 경우, 이 온톨로지들의 정확성에 대한 이슈가 추가로 발생한다. 즉, 이 온톨로지들이 과연 현재 쇼핑몰의 온톨로지와 정확히 일치하는지를 항상 확인하고 차이가 있는 경우에는 이를 반영하여야 한다. 그 외에도 수많은 쇼핑몰들의 온톨로지 보관에 대한 공간적 부담도 크다. 본 연구에서 제안하는 RSS 기반의 비교쇼핑은 각 쇼핑몰에 있는 모든 상품들을 대상으로 비교할 상품을 결정하는 것이 아니라 현재 RSS 피드로 전달된 상품들만 대상이 되므로 온톨로지 전체를 대상으로 온톨로지 매칭을 할 필요가 없다. 따라서 각 쇼핑몰들의 온톨로지를 유지하지 않고, RSS 피드에 있는 정보만으로 온톨로지 매칭을 지원하고자 한다. 이를 위해서는 RSS 피드에 충분한 정보가 있어야 하므로 첫째 이슈와도 관련이 있다.

마지막으로 온톨로지 매칭 알고리즘의 선택과 관한 이슈가 있다. 일반적인 온톨로지 매칭 알고리즘은 상품의 비교를 위해 고안된 것은 아니다. 보스턴 컨설팅 그룹의 조사에 의하면 전자상거래 환경에서 구매자들이 시도한 전체 검색 시도에서 28%는 원하는 상품을 검색하지 못했

<표 1> RSS기반 비교쇼핑의 이슈와 대안

이슈	대안	해결방안
RSS 피드의 형식과 내용	기존 RSS 피드 표준 형식을 확장. 기존 RSS 피드 표준 내에서 구현.	기존 RSS 피드 표준 내에서 구현하되 온톨로지 매칭을 위해 필요한 정보를 추가.
상품비교형식과 자체온톨로지	사용자가 상품분류 및 속성을 선택하고 이에 따라 상품을 검색. 자체온톨로지가 요구됨. 사용자가 상품 중 하나를 선택하고 유사상품을 검색. 자체 온톨로지 필요 없음.	두 비교형식을 모두 지원하고 이에 따라 자체온톨로지를 사용.
외부쇼핑몰의 온톨로지 유지	외부 쇼핑몰들의 온톨로지를 리더기 내부에 유지. 정확성 문제가 발생.	외부 쇼핑몰들의 온톨로지를 유지하지 않고 RSS 피드만으로 온톨로지 매칭을 구현.
온톨로지 매칭 알고리즘	온톨로지 매칭 효과에서 정확성, 재현율, F Measure 각각을 추구.	전자상거래 환경에 적합하도록 재현율을 높이는 매칭 알고리즘을 선택.

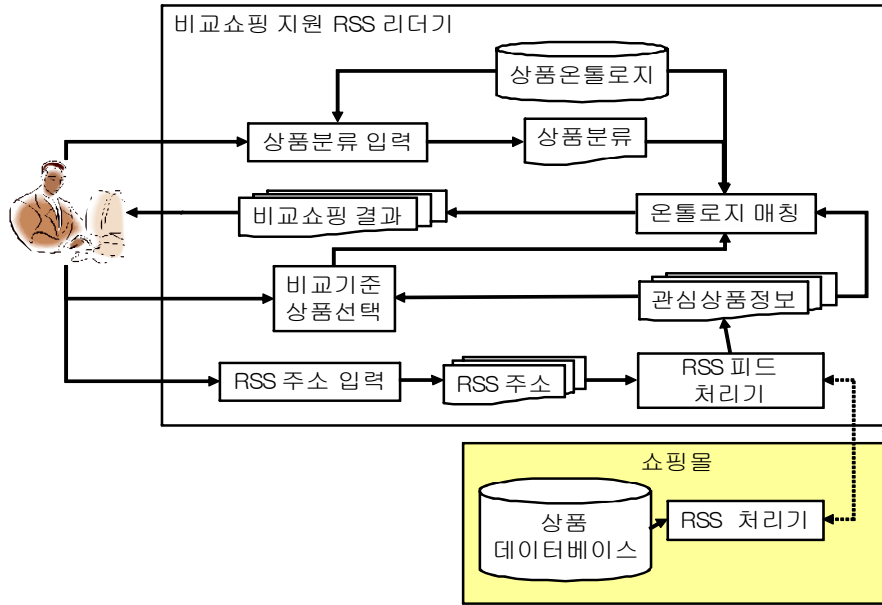
고, 구매자들의 48%는 검색결과가 만족스럽지 않았던 경험이 있는 것으로 나타났다(Pecaut et al., 2007). 온톨로지 매칭의 효과는 정확도(precision)와 재현율(recall)에 의해 평가되는데, 보수적이고 엄격한 온톨로지 매칭은 주로 정확도를 더 중요하게 보는 경향이 있다. 그러나, 이 경우에 구매자는 자신이 원하는 의도와 정확히 일치하지만 몇 개 안되는 상품 리스트를 검색 결과로 받게 될 가능성이 높고, 이는 앞서 언급한 검색의 실패로 연결될 가능성이 높다. 따라서 본 연구에서는 구매자가 원하는 상품과 유사한 상품들이 가급적 많이 검색결과에 포함되는 것을 목표로 하고자 한다. 이는 기존의 온톨로지 매칭과는 달리 적절한 선에서 정확도를 손해 보는 대신 재현율을 충분히 높임으로써 달성할 수 있다. 또한 본 논문에서는 단일 상품분류가 아닌 상품분류의 전체 경로를 이용한 온톨로지 매칭을 실시함으로써 보다 정확한 매칭 결과를 얻고자 한다. <표1>은 이상에서 논의된 이슈들과 대안들 그리고 해결방안을 요약한 표이다.

IV. 온톨로지를 이용한 RSS 기반 비교쇼핑 시스템

본 장에서는 3장에서 논의된 이슈에 대한 해결방안들을 기반으로 하여 RSS 기반 비교쇼핑의 시스템과 프로세스를 제안하고, 온톨로지 매칭 알고리즘을 설명하고자 한다.

4.1 RSS 기반 비교쇼핑 시스템 구조와 프로세스

3장에서 논의한 바와 같이 본 논문에서 제안하는 RSS 기반의 비교쇼핑은 기존 RSS 표준의 범위 내, 즉 사용자가 원하는 상품에 대한 RSS 주소를 쇼핑몰로부터 받아 RSS 리더기에 입력하고 RSS 피드를 받는 환경에서 실행된다. 제안된 이슈에 대해서는 첫째 RSS 표준을 최대한 활용하여 RSS 피드에 온톨로지 매칭을 위해 필요한 상품정보를 추가하고, 둘째 사용자가 상품분류와 속성을 선택하여 비교할 상품들을 찾는 방

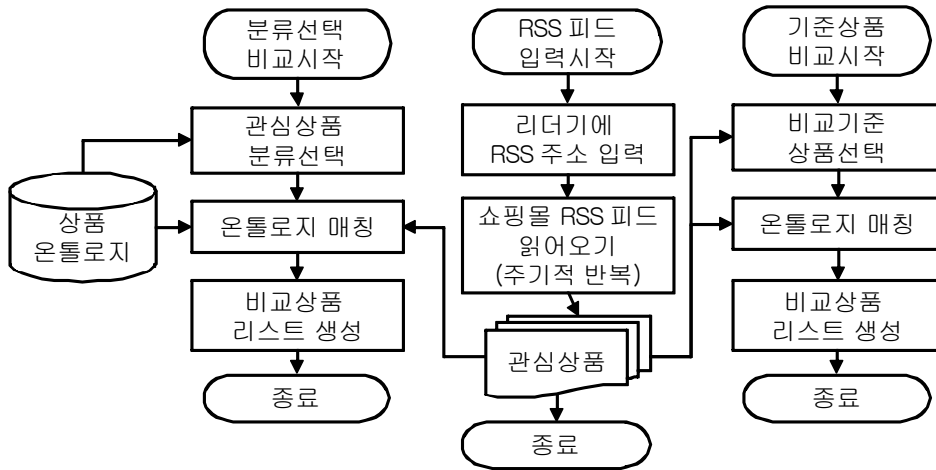


<그림 2> RSS 기반 비교쇼핑 시스템 구조

식과 하나의 상품을 선택하고 이와 유사한 상품을 검색하여 비교하는 방식을 모두 지원하며, 셋째 외부 쇼핑몰의 온톨로지를 내부에 유지하지 않도록 하였다. <그림 2>은 이를 반영하여 만들어진 RSS 기반의 비교쇼핑 시스템의 구조를 보여준다.

그림에서 사용자가 리더기에 RSS 주소를 입력하고 이 주소에 따라 쇼핑몰에 상품정보를 요청하여 가져오는 부분까지는 기존의 리더기와 동일하다. 사용자가 상품정보를 보다가 비교기준이 되는 상품을 선택하면 이 상품과 매칭되는 상품들을 온톨로지 매칭으로 관심 상품정보에서 검색하여 보여주게 된다. 사용자가 원하는 상품분류를 입력하여 상품을 비교하고자 하는 경우에는 자체적으로 유지하는 상품 온톨로지에 있는 상품분류를 이용하여 입력하도록 하고 이에 따라 관심 상품정보로부터 온톨로지 매칭을 통해 해당하는 상품들을 비교하여 보여준다.

<그림 3>은 이러한 비교쇼핑 프로세스를 보다 상세하게 보여주는 순서도이다. 그림의 가운데 부분은 사용자가 쇼핑몰에서 관심 상품에 대한 RSS 주소를 생성하고 이를 리더기에 입력하여 RSS 피드를 읽어오는 과정을 나타낸다. 이때 RSS 피드를 읽어오는 부분은 주기적으로 반복될 수 있다. 그림의 좌측은 상품의 분류를 지정하여 상품비교를 하는 과정이다. 자체 상품 온톨로지로부터 관심 상품의 분류를 지정하면, 이미 생성된 관심 상품들과의 온톨로지 매칭을 통해 비교할 상품의 리스트를 생성한다. 그림의 우측은 하나의 상품을 기준으로 하여 비교를 하는 과정이다. 관심 상품으로부터 기준이 되는 하나의 상품을 선택하여 상품비교를 실행하면, 이 상품과 관심 상품들 간의 온톨로지 매칭을 통해 마찬가지로 비교상품의 리스트를 생성한다.



<그림 3> RSS 기반 비교쇼핑 프로세스

4.2 비교쇼핑을 지원하기 위한 RSS 피드의 내용

3장에서 논의한 바와 같이 각 쇼핑몰의 온톨로지를 리더기 내에 유지하지 않기 때문에, RSS 피드의 내용은 온톨로지 매칭을 위해 필요한 상품정보들을 담고 있어야 한다. <표 2>는 현재 Amazon.com과 Buy.com에서 RSS 피드에 사용하고 있는 <item>의 하부 태그들을 보여준다. RSS 표준에서 <item>의 하부 태그들만 살펴본 이유는 RSS 표준에서 <item>이 각 상품정보들을 표현하도록 정의되어 있기 때문이다. 양쪽에

서 공통으로 사용되는 태그들은 <title>, <description>, <link>, <pubDate>로 각각 제품의 제목, 제품상세설명, 제품의 홈페이지 링크, RSS 피드 작성일자의 용도로 사용되고 있다.

Amazon.com에서는 기본 표준의 <guid>와 Atom 표준에 의한 두 개의 태그 외에 마이크로소프트에서 정의하고 있는 5개의 태그를 추가로 사용하고 있다. 반면, Buy.com은 기본 표준의 <enclosure>와 자체적으로 정의하고 있는 8개의 태그를 사용하고 있다. <그림 1>에서 살펴본 익스플로러 8에서 buy.com은 정렬기준으로 price와 rank를 제공하고 있는데, 이는 각각 <p:price>

<표 2> Amazon.com과 Buy.com의 RSS 태그사용 현황

	Amazon.com	Buy.com
공통사용태그	<title>, <description>, <link>, <pubDate>	
독립사용태그	<guid>, <atom:published>, <atom:updated>, <cfi:id>, <cfi:effectiveld>, <cfi:read>, <cfi:downloadurl>, <cfi:lastdownloadtime>	<enclosure>, <p:sku>, <p:price>, <p:shipping>, <p:shipcharge>, <p:shorttitle>, <p:listprice>, <p:rebate>, <p:pricebeforerebate>, <p:priceafterrebate>, <p:description>, <p:imagelink>, <p:rank>

와 <p:rank> 태그에 의해 제공되는 정보를 이용하고 있다. 반면, Amazon.com이 정렬기준으로 제공하는 날짜와 제목은 <title>과 <pubDate>로 예측되나, <pubDate>는 모든 상품이 동일한 값을 가지기 때문에 사실상 정렬효과가 없다.

RSS 표준에 따르면 <item>의 하부 태그인 <category>는 아이템에 해당하는 하나 이상의 분류를 나타내도록 되어 있다. 따라서 본 논문에서는 <category>를 이용하여 온톨로지 매칭에 필요한 상품의 분류를 표현하였다. 예를 들어 Amazon.com에서 LED TV는 "Electronics > Televisions & Video > Televisions > TVs > LED Lit"의 분류체계에 있다. 해당 상품분류인 LED Lit 뿐만이 아니라 상위 분류를 모두 표현하는 이유는 온톨로지 매칭에서 상품분류의 경로정보를 모두 이용함으로써 보다 정확한 매칭을 하기 위해서이다. 이 분류에 속하는 LED TV에 대한 RSS 피드의 내용은 다음과 같이 생성된다.

```
<item>
<title>LG 55LE5400 55 Inch 1080p 120 Hz
LED HDTV with Internet Applications</title>
<link>http://www.amazon.com/LG_55LE5400
55_Inch_Internet_Applications/dp/B0039JBX
VY/ref=sr_1_2?s=tv&ie=UTF8&qid=1298546383
&sr=1_2</link>
<pubDate>Thu, 24 Feb 2011 08:45:18
GMT</pubDate>
<description>&lt;div style="float:left"&gt;&lt;a
class=url ...</description>
<category domain="cat1">Electronics</category>
<category domain="cat2">Televisions &
Video</category>
<category domain="cat3">Televisions</category>
<category domain="cat4">TVs</category>
```

```
<category domain="cat5">LED Lit</category>
...
</item>
```

분류체계는 <category> 태그에서 domain 속성을 이용하여 최상위부터 최하위까지 차례로 기술한다. 이렇게 함으로써 온톨로지 매칭에 필요한 상품분류를 RSS 피드에 표현할 수 있다. 이 외에도 <category> 태그를 이용하여 상품의 비교에 필요한 속성정보를 표현할 수 있다. 예를 들어 가격정보는 <category domain="price">\$2,499.99</category>와 같이 표현 가능하다.

4.3 온톨로지 매칭 알고리즘

본 논문에서는 상품 분류에 대해 온톨로지 매칭을 수행하여 다른 쇼핑물에서 유사한 상품 분류를 찾아 매칭시키기 때문에, 온톨로지 스키마만으로 매칭을 수행하는 스키마 기반 매칭(Shvaiko and Euzenat, 2005)을 이용하여 온톨로지 매칭을 수행하였다. 기본적으로 스키마 기반 온톨로지 매칭은 한 온톨로지 내의 클래스에 대해 다른 온톨로지의 클래스들의 유사도를 계산하여 일정 기준값을 넘는 것을 매칭시키는 작업이라고 할 수 있다. 본 연구에서 제안하는 온톨로지 매칭단계는 일반적인 온톨로지 매칭과 유사하게, 워드넷(WordNet)을 이용하여 매칭하고자 하는 대상의 정확한 의미를 파악하는 단계와, 파악된 의미를 이용하여 목표 온톨로지 매칭 클래스 후보를 검색하는 단계, 그리고 후보들에 대해 유사도를 계산하여 쓰레숄드에 따라 최종 유사성을 계산하는 단계로 이루어져 있다.

첫단계인, 워드넷을 이용하여 매칭 대상의 정확한 의미(sense)를 파악하는 단계는, notebook

과 같이 노트북 컴퓨터와 공책의 두 가지 의미를 갖고 있는 상품분류가 있을 때, 상위분류를 이용하여 워드넷으로부터 이 분류의 정확한 의미를 파악하는 단계이다. 이 단계에서는 UMND1 (Unsupervised Word Sense Diambiguation Using Contextual Semantic Relatedness) (Patwardhan et al., 2007)을 이용하였다. 이 방법은 문맥(context)을 이용하여, 대상 단어의 정확한 의미를 워드넷으로부터 찾아내기 위한 방법론으로 부가적인 주석이나 감독이 필요없다는 장점이 있다. 본 논문에서 매칭 대상이 되는 상품분류는 "Electronics > Televisions & Video > Televisions > TVs > LED Lit"와 같이 상품분류의 전체 경로로서, "LED Lit"이 직접적인 상품분류라고 할 때 상위 경로의 상품분류는 UMND1에서 문맥정보로 이용할 수 있다. 매칭하고자 하는 상품분류가 t 라고 할 때, 워드넷에서 이 t 는 여러 개의 의미(sense)를 가질 수 있다. 원소의 수가 T 인 이 의미집합을 $\{t_1, t_2, \dots, t_T\}$ 라 하고, 경로에 있는 상위분류의 집합을 $\{w_1, w_2, \dots, w_n\}$ 이라고 할 때, 상위분류는 각각 W_1, W_2, \dots, W_n 개의 의미(sense)를 가지고 있을 수 있다. 이 단계의 목표는 상위분류의 집합을 고려할 때 의미집합에서 t 의 가장 그럴듯한 의미를 찾아내는 것이다. 이 때 UMND1에 따르면 t 에 대한 각 의미(sense) t_i 는 다음과 같은 적합도 점수를 갖게 된다. 식에서 w_{jk} 는 상위 분류 w_j 의 k 번째 의미(sense)를 나타낸다.

$$score(t_i) = \sum_{j=1}^n \max_{k=1 to W_j} (relatedness(t_i, w_{jk}))$$

파악된 의미를 이용하여 관심상품 리스트에서 매칭 클래스 후보를 찾는 단계는 비교적 단순하게 이루어진다. 앞 단계에서 찾아진 의미

(sense)를 이용하여 상품분류가 유사한 상품들을 검색하는데, 이 때 워드넷에서 제공하는 의미(sense)의 유사어 및 동의어를 사용하여 후보들을 찾아 저장하게 된다.

사용자가 선택한 상품분류 혹은 기준 상품분류와 관심상품의 상품분류는 앞에 설명한 바와 같이 모두 여러 개의 상품분류로 이루어진 경로이다. 따라서 두 상품분류의 유사도를 구하기 위해서는 집합 간의 유사도를 계산하여야 한다. 기준 상품분류를 $sPath = \{sP_{t_1}, sP_{t_2}, \dots, sP_{t_o}\}$, 대상이 되는 관심상품들의 상품분류들을 $\{tPath_1, tPath_2, \dots, tPath_n\}$ 으로 나타내고 하나의 관심상품 상품분류를 $tPath_i = \{tP_{t_1}, tP_{t_2}, \dots, tP_{t_m}\}$ 으로 나타낼 때, 기준 상품분류와 하나의 관심상품과의 유사도는 다음과 같이 모든 가능한 분류 조합의 문자열 유사도의 평균으로 정의하였다.

$$similarity(sPath, tPath_i) = \frac{\sum_{j=1}^o \sum_{k=1}^m label_sim(sP_{t_j}, tP_{t_k})}{o \cdot m}$$

위 식에서 두 상품분류의 유사도를 계산하는 $label_sim()$ 은 Jaro distance(Jaro, 1995)를 사용하였으며 식은 다음과 같다.

$$label_sim(s_1, s_2) = \frac{1}{3} \left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m} \right)$$

위 식에서 m 은 두 문자열 s_1 과 s_2 의 일치하는 문자의 수이며, t 는 일치하는 문자 중에서 순서가 바뀐 문자의 수이다.

이상과 같은 유사도로 모든 관심상품들의 유사도를 구한 후에, 주어진 쓰레숄드를 초과하는 유사도 값을 가진 관심상품들이 비교상품 리스트로 제공된다. 이 때, 쓰레숄드의 값을 높이면 기준이 강화되어 정확도는 높아지는 반면 재현

율은 떨어질 것으로 예상된다. 반대로 쓰레숄드의 값이 낮추면 정확도는 낮아지고 재현율은 높아질 것으로 예상된다.

V. RSS 기반 비교쇼핑 시스템의 상품매칭 성능 실험

본 논문에서 제안하는 비교쇼핑 시스템의 성능은 전적으로 온톨로지 매칭의 성능에 달려 있다고 할 수 있다. 실험은 상품 온톨로지에서 상품의 분류를 선택한 후 이와 유사한 상품들을 RSS 피드의 관련상품 리스트에서 검색하는 것과 RSS 피드의 관련상품 리스트에서 상품을 선택하고 이와 유사한 상품들을 검색하는 것으로 나누어 실시하였다.

먼저 자체 상품 온톨로지는 ODP(Open directory Project, www.dmoz.com)의 온톨로지를 이용하여 구축하였다. 사용자는 이 온톨로지로부터 비교하고자 하는 상품의 분류를 선택하게 된다. RSS 피드의 관련상품은 Amazon.com과 Buy.com 두 개의 사이트로부터 구축하였다. 상품분류는 각각 ODP로부터 146개, Amazon으로부터 151개, Buy.com으로부터 257개를 추출하였다. 상품 온톨로지로부터 비교상품을 검색하는 경우에는 ODP로부터 Amazon.com과 Buy.com으로 온톨로지 매칭이 이루어지며, 관련상품 리스트에서 상품을 선택하여 비교상품을 검색하는 경우에는 Amazon.com으로부터 Buy.com으로 그리고 Buy.com으로부터 Amazon.com으로 온톨로지 매칭이 이루어진다.

실험에서는 재현율(recall)과 정확도(precision) 그리고 F-measure 이렇게 세 개의 척

도를 사용하였다. 일반적인 분류문제에서는 정분류율(accuracy)를 기본적으로 사용하는 것이 바람직하나, 온톨로지 매칭 문제에서는 클래스 불균형 문제로 인해 재현율과 정확도가 선호되고 있다(Tan et al., 2006). 즉 하나의 상품에 대해 다른 쪽의 온톨로지에서 매칭되는 상품의 수가 전체 상품의 수에 비해 매우 적기 때문에 온톨로지 매칭 성능이 매우 낮은 알고리즘인데도 불구하고 정분류율은 높게 나올 수 있기 때문이다.

척도를 계산하기 위하여 사람이 먼저 각 상품 분류를 모두 확인하여 올바른 매칭결과를 가려냈다. pos 는 사람이 확인한 올바른 매칭결과의 수이고, t_pos 는 시스템이 제시한 매칭 결과 중에서 옳은 것의 수, f_pos 는 시스템이 제시한 매칭 결과 중에서 틀린 것의 수라 할 때, 정확도와 재현율은 다음 식과 같이 계산된다.

$$\text{정확도} = \frac{t_pos}{t_pos + f_pos},$$

$$\text{재현율} = \frac{t_pos}{pos}$$

F-measure는 정확도와 재현율의 조화평균(harmonic mean)이며 온톨로지 매칭의 전반적인 성능을 나타내는 지표로 다음과 같이 계산된다.

$$F\text{-measure} = \frac{2 \times \text{정확도} \times \text{재현율}}{\text{정확도} + \text{재현율}}$$

4장에서 설명한 온톨로지 매칭 방법론에서 쓰레숄드는 정확도와 재현율 사이의 균형을 맞추는데 사용하였다. 본 논문에서는 전자상거래 환경에 적합한 방법을 제안하고자 했기 때문에 여러 쓰레숄드에 대해 실험을 수행한 후, 적절

<표 3> 비교쇼핑 온톨로지 매칭 결과

	F measure		재현율 (%)		정확도 (%)	
	시스템	PROMPT	시스템	PROMPT	시스템	PROMPT
분류선택비교	74.9	71.0	95.2	58.9	62.0	89.4
기준상품비교	45.6	45.1	89.3	48.5	30.6	43.1
평균	60.2	58.0	92.3	53.7	46.3	66.2

한 F-measure를 유지하면서 높은 재현율을 보이는 쓰레숄드를 선택하였다. 본 실험에서는 제안된 비교쇼핑 시스템의 온톨로지 매칭 결과를 Protégé를 기반으로 한 온톨로지 매칭 프로그램인 PROMPT(Noy and Musen, 2003)의 온톨로지 매칭 결과와 비교하였다. PROMPT는 여러 개의 온톨로지를 관리하기 위한 프레임워크로서, 사용자와의 상호작용을 통해 온톨로지 병합을 수행하는 IPROMPT와 온톨로지 매칭을 위한 ANCHORPROMPT, 그리고 온톨로지 버전 관리를 위한 PROMPTDIFF 등의 도구를 제공한다. PROMPT는 서로 다른 온톨로지에 있는 개념들 간의 상호관계를 찾기 위한 그래프 기반의 알고리즘을 사용하는데, 이를 위해 클래스 이름, 속성 이름, 속성 도메인, 속성 범위 등을 사용한다. 본 논문의 알고리즘은 워드넷(WordNet)을 이용하여 매칭하고자 하는 대상의 정확한 의미를 파악하는 단계가 있다는 점에서 PROMPT와 차이가 있으며, 상품분류의 유사도를 계산하는 식에서도 차이가 있다. 무엇보다 PROMPT는 범용적인 온톨로지 매칭 도구인 것에 비해 본 논문의 알고리즘은 전자상거래를 위한 매칭 도구라는 점에서 가장 큰 차이가 있다.

<표 3>은 각 척도에 대하여 사용자가 자체 온톨로지서 상품의 분류를 선택하여 비교를

실행한 결과와 사용자가 관심상품 리스트 중에서 관심상품을 선택하여 비교를 실행한 실험결과를 보여준다. 먼저 온톨로지 매칭의 전반적인 성능을 나타내는 지표인 F-measure를 보면, PROMPT에 비해서 분류선택비교에서는 5.5% 향상되었으며, 기준상품비교에서는 1.1% 향상되었다. 이는 개별적인 상품분류를 비교하는 대신, 상품이 속한 분류의 경로를 이용하여 온톨로지 매칭을 수행했기 때문으로 판단된다. 앞서 언급한 바와 같이 전자상거래 환경에서는, 제안된 상품들이 사용자가 원하는 상품일 가능성이 높기 보다는 사용자가 원하는 상품들이 추천 리스트에 가급적 많이 포함되는 것이 더 중요하다. 그렇지 않은 경우에 검색 실패로 이어질 확률이 높기 때문이다. 각 결과에서 정확도의 차이에 비해 재현율은 PROMPT에 비해 월등하게 높은 결과를 보여 주고 있다. 전체적으로 정확도는 PROMPT에 비해 19.9%가 낮지만 재현율은 38.6%로 정확도의 손해에 비해 두 배 높아졌다. 즉 원래 목표한 전자상거래에 적합한 매칭 결과를 보여주고 있는 것을 알 수 있다.

VI. 결론

인터넷으로 상품을 구매하기 위해서는 구매

자가 여러 판매자들이 제시하는 상품들의 판매 조건과 속성에 대한 정보를 수집하고 이 정보들을 비교 평가함으로써 최선의 상품을 선택하게 되는데, 여기에는 많은 노력과 시간이 요구된다. 비교쇼핑 서비스를 제공하는 가격비교 사이트를 이용하면 가격 등의 결제조건을 포함한 상품 속성들과 상품을 제공하는 판매자들을 보다 쉽게 비교할 수 있으나, 원하는 물품이 특정 쇼핑물에 있음에도 불구하고 검색이 되지 않거나, 상품정보가 부정확한 경우가 빈번히 발생하고, 지나치게 많은 결과가 나오는 등의 단점이 있다.

이러한 비교쇼핑의 단점들을 극복하기 위해 본 논문에서는 RSS에 기반한 비교쇼핑 시스템을 제안하였다. RSS는 각 쇼핑물에서 직접 제공되기 때문에 완전성과 부정확성의 문제를 해결하고 사용자가 관심있는 상품분류만을 제공함으로써 개인화된 결과를 제공할 수 있다. 이전 연구에서 제안된 지능형 RSS 프레임워크는 개별적인 쇼핑물들에 대한 방문 없이 지능형 서비스를 이용해 한번에 원하는 상품들을 비교할 수 있으나 모든 사이트가 광범위한 표준을 따라야 한다는 제약조건이 있었다. 본 논문에서는 주어진 표준 범위 내에서 RSS 리더기 수준의 개인화된 비교쇼핑결과를 제공하기 위한 시스템을 제시하였다.

표준화되지 않은 상품분류 및 속성으로 각 쇼핑물이 보내오는 RSS 피드를 이용하여 비교쇼핑을 지원하기 위해 본 논문에서는 온톨로지 매칭을 이용하여 의미적 상호운용성을 구현하였다. 온톨로지 매칭은 상품분류에 대하여 워드넷에서 정확한 의미(sense)를 찾아내기 위한 단계, 파악된 의미를 이용하여 목표 온톨로지에서 매칭 클래스 후보를 검색하는 단계, 그리고 후보

들에 대해 유사도를 계산하여 쓰레숄드에 따라 최종 유사성을 계산하는 단계로 구현하였다. 구현한 비교쇼핑 시스템을 이용하여 범용 온톨로지 매칭 도구인 PROMPT와 비교하여 실험한 결과, F-measure에서 약간의 향상이 있었으며 원래의 목표인 재현율에서는 38.6%가 더 높았다. 정확도에서 19.9% 더 낮은 값을 보였으나 사용자에게 원하는 상품과 유사한 상품들을 가급적 많이 추천해야 하는 전자상거래 환경에서는 보다 효과적이며, 이는 온톨로지 매칭을 이용한 RSS 기반의 비교쇼핑이 효과적임을 보여준다.

본 논문에서 제안하는 RSS 기반의 비교쇼핑을 이용할 경우, 기존의 수작업에 의한 비교쇼핑 사이트의 업데이트를 벗어나 자동화된 방법으로 새로운 상품들을 비교할 수 있으며, 정확성과 완전성 또한 향상될 수 있을 것으로 기대된다. 뿐만 아니라, 사용자가 관심있는 판매자와 상품분류에 대해서만 비교를 수행함으로써 보다 유용하고 개인화된 비교쇼핑 서비스를 제공할 수 있을 것으로 기대된다.

참고문헌

- 권이남, 김재수, “학회정보 개인화 지원을 위한 RSS기반 유통 모델 연구,” 한국콘텐츠학회 추계종합학술대회, 제2권, 제2호, 2004, pp.545-553.
- 권이남, 김재수, 신동구, 전성진, “RSS기반 과학 기술정보 배급 표준시스템 구현에 관한 연구,” 한국콘텐츠학회 추계종합학술대회, 제3권, 제2호, 2005, pp.60-65.

- 김성언, “인터넷 쇼핑물 구매경험자들의 고객만족도에 영향을 미치는 요인,” 정보시스템연구, 제17권 제2호, 2008, pp.27-47.
- 노상규, 안정남, “비교쇼핑을 위한 판매자 추천 방법에 관한 연구,” Information Systems Review, 제9권, 제2호, 2007, pp.109-127.
- 박상언, 강주영, 김우주, “전자상거래의 지능형 RSS 도입을 위한 이슈 분석과 지능형 RSS 프레임워크의 제안,” 한국전자거래학회지, 제12권, 제2호, 2007, pp.269-290.
- 이기성, 유영훈, 조근식, “시맨틱 웹 기반의 이미지 검색을 이용한 비교 쇼핑 시스템,” 한국지능정보시스템학회논문지, 제11권, 제2호, 2005, pp.1-15.
- 이머징 테크, “RSS마케팅/광고의 특징과 그 가능성,” http://www.etechtrend.co.kr/2006/05/rssaeaiac_au_e.html, 2006.
- 이한진, 염창선, “구매자 선호도를 고려한 비교 쇼핑 에이전트 시스템의 개발,” 산업경영시스템학회지, 제28권, 제1호, 2005, pp.84-91.
- 전종홍, 이승윤, “웹 2.0 기술 현황 및 전망,” 전자통신동향분석, 제21권, 제5호, 2006, pp.141-153.
- 주재훈, “시맨틱 웹 기반의 개인화 지능형 문화관광 서비스 시스템,” 정보시스템연구, 제18권, 제3호, 2009, pp.211-229.
- 중앙일보, “인터넷가격비교사이트에누리닷컴, 가장믿을만,” http://pdf.joinsmsn.com/article/pdf_article_prv.asp?id=DY01200812180163, 2008년 12월 18일.
- 채혁기, 박상언, 강주영, “인터넷 쇼핑물을 위한 AISAS 모델 기반의 RSS 마케팅 활용방안에 관한 연구,” 한국전자거래학회지, 제13권, 제3호, 2008, pp.21-49.
- 한영춘, 서영석, “Web2.0 기반 인터넷쇼핑물의 구현,” 인터넷전자상거래연구, 제9권, 제3호, 2009, pp.49-67.
- Benetti, H., Beneventano, D., Bergamaschi, S., Guerra, F., and Vincini, M., "SI Designer: An Information Integration Framework for e Commerce," *IEEE Intelligent Systems*, Vol. 17, No.1, 2002, pp. 8-25.
- Blansit, B. D., "Using RSS to Publish Library News and Information," *Journal of Electronic Resources in Medical Libraries*, Vol.3, No.1, 2006, pp.97-104.
- Chang, Y. S., Lee, and K. J., "A Comparison Shopping Optimization Model Based on Suppliers' Pricing Contexts," *Expert Systems with Applications*, Vol.37, 2010, pp.5736-5744.
- Cheung, K., Yip, K. Y., Smith, A., deKnikker, R., Masiar, A., and Gerstein, M., "YeastHub: A Semantic Web Use Case for Integrating Data in the Life Sciences Domain," *Bioinformatics*, Vol.21, SUP 1, 2005, pp.85-96.
- Curran, K., McKinney, S., "Delivering Audio and Video with Rich Site Summary Enclosures," *Journal of Multimedia*, Vol.1, No.1, April 2006, pp.24-31.
- Ehrig, M., and Staab, S., "QOM: Quick

- Ontology Mapping," Lecture Notes in Computer Science, No.3298, 2004, pp.683 - 697.
- Garfinkel, R., Gopal, R., Parhak, B., F., and Yin, "Shopbot 2.0: Integrating Recommendations and Promotions with Comparison Shopping," *Decision Support Systems*, Vol.46, 2008, pp. 61 - 69.
- Jaro, M. A., "Probabilistic Linkage of Large Public Health Data File," *Statistics in Medicine*, Vol.14, 1995, pp.491 - 498.
- Karger, D.R., Quan, D., "What Would It Mean to Blog on the Semantic Web?," *Journal of Web Semantics*, Vol.3, No.2/3, 2005, pp.147-157.
- Miller, G.A., "WordNet: A lexical Database for English," *Communications of the ACM*, Vol.38, No.11, 1995, pp.39 - 41.
- Noy, N.F. and Musen, M.A. "The PROMPT Suite: Interactive Tools for Ontology Merging and Mapping," *International Journal of Human Computer Studies*, Vol.59, No. 6, 2003, pp.983 - 1024.
- O'Reilly, T., "What is Web 2.0", http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what_is_Web_20.html, 2005.
- Patwardhan, S., Banerjee, S. and Pedersen, T., "UMND1: Unsupervised Word Sense Disambiguation Using Contextual Semantic Relatedness," *Appears in the Proceedings of SemEval 2007: 4th International Workshop on Semantic Evaluations*, June 2007, pp.390-393.
- Pecaut, D., Silverstein, M. and Stanger, P., "Winning the Online Consumer in Canada: Insights into Online Consumer Behavior," Boston Consulting Group Report, 2007.
- Porter, M. E., "Consumer behavior, Retailer Power and Market Performance in Consumer Goods Industries," *The Review of Economics and Statistics*, Vol.56, No.4, Nov. 1974, pp.419-436.
- Rethlefsen, M. L., Segovis, C. M., "Cool Tools: Google, RSS, Podcasts, Oh My! A look at Some Applications that can Make it Easier to Get Up to Date Medical Information," *Minnesota Medicine*, Vol.89, No.11, 2006, pp. 32-36.
- Shvaiko, P., Euzenat, J., "A Survey of Schema Based Matching Approaches," *Journal on Data Semantics*, No.3730, 2005, pp.146-171.
- Staat, M., H. H. Bauer, and M. Hammerschmidt, "Structuring Product Markets: An Approach Based on Customer Value," *American Marketing Association*, Winter 2002, pp. 205-212.
- Talluri, S., R. Narasimhan, and A. Nair, "Vendor Evaluation with Supply Risk: A Chance Constrained DEA Approach," *International Journal of Production Economics*, Vol.100, 2006. pp.212-222.
- Tan, P. N., Steinbach, M., Kumar, V.,

"Introduction to Data Mining," *Addison Wesley*, 2006. pp.294-298.

Veltman, K. H., "Syntactic and Semantic Interoperability: New Approaches to Knowledge and the Semantic Web," *New Review of Information Networking*, Vol.7, 2001, pp.159-184.

Zhu, J., "A Buyer Seller Game Model for Selection and Negotiation of Purchasing Bids: Extensions and New Model," *European Journal of Operational Research*, Vol.154, 2004, pp.150-156.

박상언(Park, Sang-Un)



한국과학기술원에서 전산학과 학사, 경영공학과 석사 및 박사학위를 취득하였으며, 현재 경기대학교 경상대학 경영정보학과에 재직 중이다. 주요관심분야는 웹기반 지능

정보시스템, 시맨틱 웹 마이닝, 시맨틱 웹 검색, 온톨로지 매칭, 지능형 웹서비스, 지능형 전자상거래이다.

<Abstract>

Comparison Shopping System Based on RSS with Ontology Matching

Park, Sang-Un

In order to buy products through the Internet, consumers dissipate much time and efforts in collecting and comparing product information from various online shopping malls. Consumers can save their efforts by using price comparison sites, but there are some shortcomings in comparison shopping. Firstly, comparison sites do not show the lowest price of some products that are selling in shopping malls. Secondly, the product information provided by comparison sites is sometimes wrong. Thirdly, there are too many results. In order to overcome the shortcomings, we suggested a comparison shopping system based on RSS by using ontology matching. We used the current RSS standard for syntactic interoperability instead of suggesting new standards. Moreover, we used ontology matching for semantic interoperability to compare product information with different ontologies. The suggested ontology matching consists of three steps. The first step is finding exact sense from WordNet for a given product category, and the second step is searching for matching product category candidates from the products of RSS feeds. The final step is calculating similarities of the candidates with the target product category. From the experiments, we could get better recall rates that are suitable for e-commerce environments and the results show that our system is effective in product comparison.

Key Words : RSS, Web 2.0, Semantic Web, Comparison Shopping, Ontology Matching

* 이 논문은 2011년 3월 4일 접수되어 1차수정(2011년 4월 5일)과 2차수정(2011년 5월 17일), 3차수정(2011년 5월 21일)을 거쳐 2011년 8월 26일 게재 확정되었습니다.