

구매자 카테고리 기반 지능형 e-Commerce 메타 서치 엔진

김경필 · 이상훈 · 김창욱[†]

연세대학교 정보산업공학과

Buyer Category-Based Intelligent e-Commerce Meta-Search Engine

Kyung Pil Kim · Sang Hoon Woo · Chang Ouk Kim

Department of Information and Industrial Engineering, Yonsei University, Seoul, 120-749

In this paper, we propose an intelligent e-commerce meta-search engine which integrates distributed e-commerce sites and provides a unified search to the sites. The meta-search engine performs the following functions: (1) the user is able to create a category-based user query, (2) by using the WordNet, the query is semantically refined for increasing search accuracy, and (3) the meta-search engine recommends an e-commerce site which has the closest product information to the user's search intention by matching the user query with the product catalogs in the e-commerce sites linked to the meta-search engine. An experiment shows that the performance of our model is better than that of general keyword-based search.

Keyword: e-Commerce, Meta-Search Engine, Category based search, Semantic search

1. 서론

인터넷과 World Wide Web은 초기에는 정보 제공의 목적으로만 사용되었지만, 현재는 전자상거래라는 이윤창출의 목적으로 사용되고 있다. 전자상거래는 시간과 공간을 초월하여 구매자와 공급자의 상거래가 인터넷상에서 수행될 수 있도록 도와주는 역할을 수행한다. 초기 전자상거래 시장은 일반 구매자를 대상으로 하는 쇼핑몰과 같은 B2C 전자 상거래를 통해 인식되기 시작하였으며, 전자 상거래의 효용성이 증명됨에 따라 점차 기업을 대상으로 하는 e-marketplace와 같은 B2B 전자상거래의 비중이 급격히 높아지고 있다. 실제로 UN의 한 보고서에 따르면 전세계 B2B 전자상거래 시장이 해마다 급격하게 성장할 것이라고 예측했다(UN, 2003).

전자상거래 사이트가 기하급수적으로 증가함에 따라 다수의 공급자 중에서 구매자가 원하는 제품 정보를 가진 공급자를 찾는 것이 중요하게 인식되고 있다. 현재는 공급자 검색 시 인터넷 검색 엔진을 통하여 전자 상거래 사이트를 검색하는

방법이 사용되고 있다. 하지만, 구매자가 인터넷 검색 엔진을 통하여 다양한 전자상거래 사이트에서 원하는 조건에 맞는 제품을 찾기 위해서는 각 공급자가 제공하는 제품 정보를 구매자가 직접 수집하고 각각의 정보를 비교·분석한 후 공급자를 선택하는 과정이 요구되기 때문에 많은 시간이 소요된다는 단점이 있다.

인터넷 검색의 효율성을 증대하기 위해서 현재까지 MetaCrawler(Selberg *et al.*, 1997), Savvysearch(A. E. Howe *et al.*, 1997) NECI Metasearch Engine(S. Lawrence *et al.*, 1998) Copernic, Dogpile과 같은 한 번의 질의로 다수의 검색엔진에 동시에 질의를 할 수 있는 메타 검색 엔진(meta-search engine)이 개발되었다. MetaCrawler와 NECI Metasearch Engine은 다수의 검색 엔진의 질의 결과인 링크 페이지들의 문맥적인(contextual) 분석을 통해 이용할 수 없는 페이지의 링크는 제외하고, 관계있는 결과는 반환하여 검색 결과에 대한 효율성 및 정확성을 높이고자 하였다. Savvysearch는 최근 검색 결과의 정확도를 측정 방법론을 도입하여 측정하고 이 후의 검색 시에 정확도가 높은 순서대

[†]연락처 : 김창욱 교수, 120-749 서울시 서대문구 신촌동 134 연세대학교 정보산업공학과, Fax : 02-364-7807, Email : kimco@yonsei.ac.kr
2006년 6월 접수, 1회 수정 후 2006년 8월 게재확정.

로 링크 페이지를 반환하여 검색 정확성을 높이고자 하였다. 하지만 이들 메타 검색 엔진은 인터넷 페이지를 검색하는 다수의 일반 검색 엔진을 통합한 형태이고 키워드 기반의 검색에서 벗어나지 못하고 있어 전자 카탈로그 검색을 통한 제품 정보를 수집하기 위한 메타 검색 엔진으로 사용하기 위해서는 효율적이지 못하다.

검색의 정확성과 사용자의 편의성을 높이기 위하여 개인화(personalization)에 대한 연구(F. Liu *et al.*, 2002)가 진행되고 있다. 이 연구는 구매자가 이전에 방문한 페이지를 분석하여 주요 단어를 로그 형식으로 기록하고 차후에 검색을 할 때 이 단어들을 이용하여 검색 질의를 자동적으로 재구성한다. 이를 통하여 사용자의 검색 결과에 대한 만족도를 높이고자 하였다. 검색의 개인화에 대한 연구는 지능형 검색이라는 새로운 분야를 제시했다는 공헌도가 있다. 그러나 이 연구의 문제점은 구매자가 검색할 때 이전과는 전혀 다른 의도로 검색할 수 있는 상황에서는 정확한 검색이 이루어질 수 없다는 것이다.

구매자의 검색 의도를 정확하게 반영하기 위한 연구로 OntoSeek(N. Guarino *et al.*, 1999)이 있었다. 이 연구에서는 구매자의 질의를 의미적으로 풍부하게 표현할 수 있도록 검색 용어간의 관계(relation)를 그래프 형식으로 표현하여 검색 의도를 명확하게 하고자 하였다. 또한 전자 카탈로그의 다양한 용어 사용으로 인한 구매자 질의어와 전자 카탈로그의 제품 표현에 대한 용어 일치 문제를 WordNet(G. A. Miller, 1995)을 이용하여 해결하고자 하였다. WordNet은 인터넷상에 존재하는 일종의 영어 단어 사전(a thesaurus of English words)으로서 특정 단어의 유의어(synonym), 반의어(antonym), 상위어(hypernym), 하위어(hyponym)를 보여주는 기능을 한다. 그러나 OntoSeek은 메타 검색이 아닌 단일 공급자 전자 카탈로그 검색 도구이며, OntoSeek을 메타 검색 엔진으로 사용하려면 계층적 제품 분류 체계와 제품 속성 정보로 구성되어 있는 모든 공급자 전자 카탈로그가 표준화되어 있어야 한다. OntoSeek에서는 전자 카탈로그의 표준화를 위한 방법으로 전자 카탈로그를 온톨로지(ontology)로 만드는 것을 제안하였다.

온톨로지란 웹에서 정보를 교환하는 응용프로그램들 간에 상호운용을 제공하기 위해 용어와 용어들 간의 관계를 의미적이고 계층적으로 표현한 추상구문이며, 이를 기계가 해석할 수 있는 언어로 표현한 것이다(M. Uschod *et al.*, 1996; B. Chandrasekaran *et al.*, 1999). 온톨로지를 표현하는 대표적인 언어로는 W3C와 ISO에서 각각 제정한 OWL(Web Ontology Language)과 XTM(Xml TopicMap)이 있다. 전자 카탈로그의 제품 정보를 온톨로지로 표현하는 것은 e-commerce 시스템의 통합 검색 기능을 향상시킬 것이며, 또한 전자상거래의 서비스 품질과 효율성을 향상시킬 것이다. (H. Lee *et al.*, 2005). 현재 이질적인 전자 카탈로그 표준화를 위한 온톨로지 연구는 비즈니스 분야의 온라인 제품 검색을 통일적으로 처리하기 위한 기관인 Dublin Core, UNSPSC 등에서 이루어지고 있고, extended entity relationship 모델(H. Lee *et al.*, 2005)을 이용하여 전자 카탈로그를 표현하는 연구도 있다.

하지만 다양한 전자상거래 사이트가 가지고 있는 이질적인 전자 카탈로그를 통합하는 과정에서 자동화가 이루어지지 않았기 때문에 부가적인 유지·보수비용이 필요할 뿐만 아니라 시간이 많이 걸리는 문제가 있다. 또한 새로운 온톨로지가 나타나면 서로 간에 다시 통합해야 하는 문제점도 발생한다. 이런 이유 때문에 현재까지 제품 정보를 온톨로지로 표현하여 구현되어 있는 실제 사례가 거의 없는 실정이다(J. Shim *et al.*, 2006).

본 연구에서는 분산되어 있는 공급자의 전자 카탈로그를 통합하지 않고 제품을 검색하는 메타 검색 엔진을 제안한다. 이 검색 엔진에서는 (1) 사용자의 제품 검색 의도를 정확하게 표현하기 위해서 키워드 기반 질의가 아닌 superclass/subclass 형태의 카테고리 기반 질의 기능을 제공하며, (2) 카테고리 기반의 질의를 각 공급자의 제품 카탈로그와 매칭하여 찾고자 하는 제품을 제공하는 공급자를 선택한다. 따라서 본 논문에서는 다양한 공급자의 전자상거래 사이트에서 제공되는 서로 다른 전자 카탈로그를 온톨로지로 통합하는 관점이 아닌, 구매자가 작성한 카테고리 기반 질의와 공급자가 제공하는 전자 카탈로그를 매칭(matching)하는 관점에서 제품 검색을 바라보고자 한다.

본 연구는 다음과 같이 구성된다. 2장에서는 제안된 메타 검색 엔진의 구성 모듈과 카테고리 기반의 의미 검색 절차를 소개한다. 3장에서는 구매자의 카테고리 기반 질의의 생성 절차와 이 때 사용하는 용어를 자세히 설명한다. 4장에서는 구매자의 카테고리 기반 질의와 공급자의 제품 카탈로그를 매칭할 때 발생하는 매칭 종류를 예제와 함께 소개하고, 매칭 알고리즘을 설명하며, 매칭 결과에 따른 공급자 우선 순위를 결정하는 의미적 연관성 척도를 수리적으로 정의한다. 5장에서는 네 종류의 공급자 제품 카탈로그를 이용하여 카테고리 기반의 의미 검색의 정확성을 실험한 내용 및 분석 결과를 소개한다. 마지막으로 5장에서는 본 연구의 결론을 맺고 추후 연구 과제에 대해서 논의한다.

2. 카테고리 기반 의미 검색 절차

공급자의 전자 카탈로그(e-catalog)는 일반적으로 제품의 분류 체계인 제품 카테고리(product category)와 카테고리 노드에 존재하는 제품 속성(product attributes) 정보로 구성되어 있다. 본 연구에서는 공급자들이 HTML, XML, OWL 등 웹 문서 언어로 제품 카테고리를 표현하고, 이를 메타 검색 엔진에 웹 서비스 등 EDI방법을 통해서 제공한다는 가정을 한다. 이런 가정은 서론에서 언급하였듯이 모든 공급자들의 제품 카테고리화 제품들의 속성을 표준화하는 전자 카탈로그 통일화 작업보다 훨씬 수월하며 현실적으로도 가능한 접근 방법이다. 메타 검색 엔진 내에서는 공급자 제품 카테고리를 JENA API를 통해 트리 형태의 자료 구조로 변환하여 보관한다. 구매자는 메타 검색 엔진에서 제공하는 사용자 인터페이스를 이용해 카테고리 기반의 질의를 생성하며, 이 또한 내부적으로 트리 형태의 자료 구조로 표현

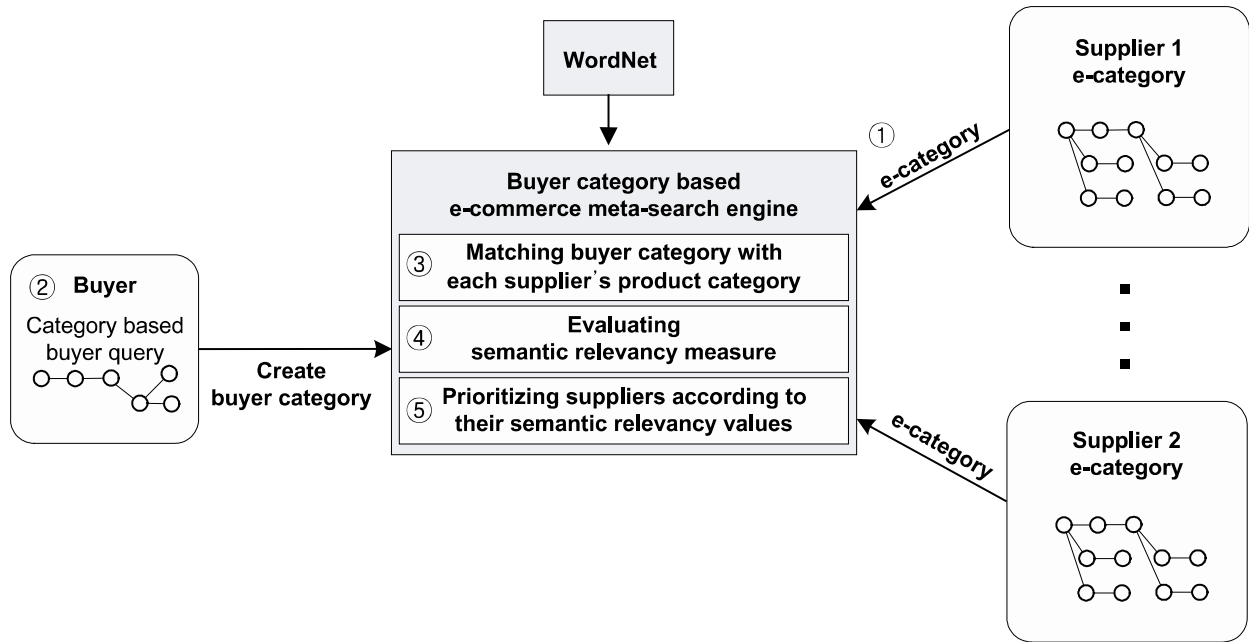


그림 1. Buyer category-based e-Commerce meta-search engine.

된다. 앞으로는 구매자의 카테고리 기반 질의를 편의상 구매자 카테고리로 부르기로 한다.

본 연구에서 제안하는 메타 검색 엔진은 <그림 1>과 같은 모듈로 구성되어 있다. 메타 검색 엔진에 연결된 각 공급자는 제품 카테고리를 OWL, XML, HTML과 같은 웹 언어로 메타 검색 엔진에 제공해 준다는 가정을 한다(<그림 1>의 ①). 구매자 카테고리를 처리하는 순서는 다음과 같다.

- Step 1. 구매자는 카테고리를 생성하며, 이 때 질의 의도를 좀 더 명확하게 하기 위해서 참조 노드를 달 수 있다 (<그림 1>의 ②).
- Step 2. 구매자 카테고리와 공급자 제품 카테고리와의 매칭 알고리즘을 통해서, 공급자 제품 카테고리의 매칭된 부분인 제품 서브 카테고리(product sub-category)를 찾아낸다(<그림 1>의 ③).
- Step 3. 매칭된 공급자의 서브 제품 카테고리 and 구매자 카테고리간의 의미적 연관성 척도 (semantic relevancy measure) 값을 측정한다(<그림 1>의 ④).
- Step 4. 모든 공급자 제품 카테고리의 매칭 작업이 끝나면 Step 5로 가고, 그렇지 않으면 Step 2와 3을 공급자 별로 반복 수행한다.
- Step 5. 측정된 의미적 연관성 척도 값이 큰 제품 서브 카테고리 순서로 구매자에게 공급자를 추천하며 (<그림 1>의 ⑤), 구매자는 이 순서대로 공급자의 홈페이지에 가서 제품 정보를 열람한다.

이 방법은 수많은 공급자가 있을 경우 구매자는 찾고자 하는 제품과 유사한 제품을 가지고 있는 공급자 순으로 공급자

제품 카테고리를 열람할 수 있는 장점이 있다. 키워드 기반의 (메타) 검색 엔진은 공급자 제품 카테고리 상에 있는 제품 계층을 단순히 AND나 OR의 조합으로만 표현할 수 있기 때문에 구매자 카테고리처럼 제품 계층을 질의어에 포함할 수 없는 한계가 있다. 따라서 키워드 기반의 검색은 카테고리 기반 (메타) 검색 엔진보다 검색 정확도가 떨어진다. 이는 5장의 실험 부분에서 확인할 수 있다.

Step 3에서 명시한 의미적 연관성 척도는 언어적 척도(linguistic measure)와 구조적 척도(structural measure)로 구성되어 있으며, 전자는 구매자 카테고리와 공급자의 매칭된 제품 서브 카테고리에 있는 용어간의 언어적 유사도를 계량화 한 것을 의미하며, 후자는 두 카테고리 트리 구조간의 유사도를 계량화 한 것이다.

다음 절부터는 Step 1에서 언급한 구매자 카테고리 생성에 대해서 자세히 설명하겠으며, 구매자 카테고리와 공급자 제품 카테고리와의 매칭 알고리즘과 의미적 유사도에 대해서는 4장에서 수리적으로 정형화하여 자세히 소개한다.

3. 구매자 카테고리 생성

본 연구에서 제안하는 카테고리 기반 질의에서는 구매자가 검색 의도를 superclass/subclass 관계인 트리 구조로 표현한다. <그림 3>는 'computer' 부속품인 'memory'를 찾고자 하는 구매자 카테고리를 보여준다. 이 예에서 검색 의도를 표현한 모든 노드들을 검색 노드(search node)라 하고, 실질적으로 찾고자하는 제품이 있는 맨 하단 검색 노드를 목표 노드(target node)라 부른다. <그림 2>에서 검색노드는 'product', 'electronic device', 'memory'에 해당하며, 목표 노드는 'memory'에 해당한다.

카테고리 기반 구매자 질의어를 사용하여 검색하는 경우에도 구매자의 모호한 카테고리 생성이 다음의 Case 1과 2와 같이 검색의 비효율을 가져올 수 있다.

(1) Case1 - <그림 2>에서 보듯이 구매자가 제품의 의미를 모호하게 카테고리를 생성했을 때

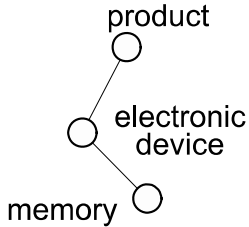


그림 2. 구매자 카테고리형 질의.

위의 카테고리는 구매자가 'memory'의 의미를 모호하게 작성했기 때문에 구매자가 원하는 제품이 컴퓨터의 부품인 'memory'인지, 휴대용 전화기에 들어가는 'memory'인지 구분할 수 없다. 이런 문제를 해결하기 위해 구매자가 카테고리를 생성할 때 참조 노드(reference node)를 생성할 수 있도록 한다. 참조 노드는 목표 노드의 형제 노드(sibling node)로 정의되며, 목표 노드의 의미를 명확히 하고자할 때 사용한다. <그림 3>에서 CPU는 목표 노드 'memory'의 참조 노드이다.

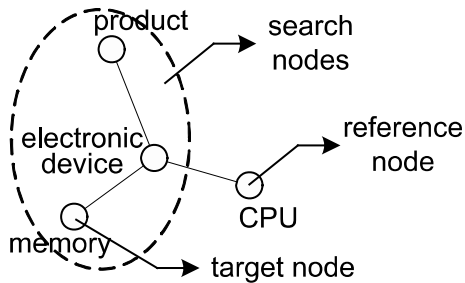


그림 3. 검색 노드, 목표 노드 및 참조 노드.

(2) Case2 - <그림 4>에서 보듯이 구매자가 생성한 카테고리 와 제품 카테고리의 의미는 같지만 용어가 다를 때

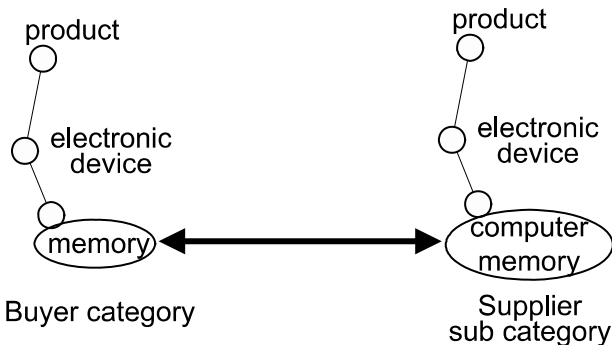


그림 4. 구매자 카테고리 와 매칭된 공급자 서브 카테고리.

이 문제를 해결하기 위해 본 시스템은 WordNet에서 구매자가 결정한 검색 노드의 유의어를 이용하여 구매자가 찾기 원하는 용어뿐만 아니라 유의어도 공급자 제품 전자 카탈로그에서 검색 함으로써 검색의 정확성을 향상시킨다. 즉 <그림 4>에서와 같이 구매자 카테고리의 목표 노드인 'memory'에 해당하는 용어를 공급자 카테고리에서 찾을 때 유의어인 'computer memory'도 매칭된 제품명으로 간주한다.

다음 장에서는 구매자 카테고리의 생성이 완료된 후 공급자 카테고리에서 구매자 카테고리 와 유사한 부분을 찾아내는 카테고리 매칭 알고리즘을 설명하겠다.

4. 전자 카테고리 매칭 알고리즘 및 의미적 연관성 척도

4.1 매칭 종류

본 논문에서 제안하는 메타 검색 엔진은 앞 장에서 언급했듯 WordNet을 이용하여 구매자 카테고리를 확장한 후, 각 공급자 제품 카테고리 와 매칭시킨다. 공급자의 제품 카테고리는 해당 공급자의 모든 제품에 대한 superclass/subclass 정보를 표현한 것이기 때문에 많은 노드를 가지고 있으나 의미적 연관성 척도를 평가하기 위해서는 구매자 카테고리 와 비교하여 제품 카테고리의 일부를 분리해서 평가해야 한다. 구매자 카테고리 와 분리된 공급자 서브 카테고리가 매칭 알고리즘의 아웃풋이며, 매칭된 결과는 다음과 같이 네 종류로 구분할 수 있다.

- Exact matching : <그림 5>의 매칭된 서브 카테고리 1처럼 구매자 카테고리 와 공급자 제품 카테고리의 노드에서 사용되는 용어 (상품명)와 구조 모두가 정확하게 일치할 경우이다.
- Non-exact matching : 구매자 카테고리 와 공급자 제품 카테고리가 exact matching이 아닐 경우를 말하며, 다시 다음의 3가지 경우로 나뉜다.
 - Synonym based exact matching : 구매자 카테고리의 노드들과 공급자 제품 카테고리의 노드들의 용어가 일부 일치하고, 나머지 모두는 WordNet에서 정의한 유의어와 일치하며, 구조 또한 정확히 일치할 경우이다. <그림 5>의 서브 카테고리 2의 경우, 트리 구조는 구매자 카테고리 와 정확히 일치한다. 또한 용어 중에서 {electronic device, computer hardware}는 정확히 일치하며, {computer memory, central processing unit}은 구매자 카테고리의 {memory, CPU}와 유의어 관계이다.
 - Original based non-exact matching : 구매자 카테고리의 노드들과 공급자 제품 카테고리의 노드들의 용어와 구조가 일부 일치하며, 나머지 용어들은 유의어를 사용하더라도

일치하는 않는 경우이다. <그림 5>의 매칭된 서브 카테고리 3의 경우, 용어중의 {electronic device, memory}는 일치하지만 나머지 용어들은 WordNet을 이용하더라도 유사하지 않으며, 구조 또한 electronic device를 root 노드로 갖고 memory를 leaf 노드로 갖는다는 것을 제외하고는 구매자 카테고리의 구조와 상이하다. <그림 5>의 서브 카테고리 5의 경우에는 ‘memory’를 제외하고는 용어와 구조가 모두 일치하지 않는 경우이다.

- Synonym based non-exact matching : 용어간의 관계에서는 synonym based exact matching과 같지만, 구조가 일부 일치할 경우이다. <그림 5>의 서브 카테고리 4에서 {computer, computer memory, CPU}는 정확히 일치하거나 유의어 관계이지만, 트리 구조는 구매자 카테고리라 일부 일치한다.

만일 공급자 제품 카테고리가 매칭 시 위의 네 가지 중 어느 것에도 속하지 않으면 검색 대상에서 제외된다. 매칭 후에, 각각의 공급자 제품 전자 카테고리는 구매자가 작성한 카테고리와의 연관성을 의미적 연관성 척도(semantic relevancy measure)를 통해 평가 받게 된다. 만일 공급자 제품 카테고리에서 매칭된 서브 카테고리가 한 개 이상 나올 경우, 서브 카테고리 중에서 의미적 연관성 척도가 가장 높은 서브 카테고리만 검색 대상으로 채택한다. 예를 들어, <그림 5>의 카테고리 4의 경우, {computer, computer memory, CPU}외에 {memory}도 단일 노드(single node)를 갖는 서브 카테고리인데 의미적 연관성 척도 값이 낮아서 고려 대상에서 제외된다.

다음 절에서는 매칭 알고리즘과 의미적 연관성 척도에 대해 알아보겠다.

4.2 매칭 알고리즘

공급자 제품 카테고리는 해당 공급자의 모든 제품에 대한 정보를 표현한 것이기 때문에 일반적으로 많은 수의 노드로 구성되나, 의미적 연관성 척도를 평가하기 위해서는 구매자 카테고리와 비교하여 이와 매칭되는 제품 카테고리의 일부분을 따로 분리해서 평가하는 것이 바람직하다. 이를 위해서 본 절에서는 구매자 카테고리라 매칭되는 제품 카테고리의 일부를 분리하기 위한 제품 카테고리 매칭 알고리즘을 제안한다.

제안하는 매칭 알고리즘을 설명하기에 앞서 인덱스 i 를 사용하여 먼저 구매자 카테고리의 검색 노드에 속하는 최하위 노드부터 상위 노드의 순으로 번호를 부여하기로 한다. 또한 검색 노드의 수를 n 으로 나타내기로 한다. 예를 들어 <그림 4>의 구매자 카테고리의 경우, ‘memory’를 표현하고 있는 목표노드는 $i=1$ 인 노드가 되며, ‘computer hardware’와 ‘electronic device’는 각각 $i=2, i=3$ 이며 $n=3$ 이 된다. 또한 이와 별도로 참조 노드의 경우 $i=0$ 으로 설정하기로 한다. 구매자 카테고리를 기준으로 하여, 매칭 알고리즘은 목표 노드인 $i=1$ 에서부터 시작하여

i 를 순차적으로 증가시키면서 i 와 일치하거나 WordNet에서 정의한 유의어 관계가 성립하는 노드를 공급자 제품 카테고리에 찾는다. 이러한 매칭 관계가 성립하는 노드를 선별하여 이들을 연결하는 sub-tree를 구성하면 공급자 제품 카테고리의 분리는 완성된다. 매칭 알고리즘은 공급자마다 적용하게 되며, 구매자 카테고리라 공급자 제품 카테고리의 쌍이 입력이 되고, 매칭된 공급자 서브 카테고리가 출력이 된다. 매칭 알고리즘은 다음과 같다.

제품 카테고리 매칭 알고리즘

Step 0: $i=0, P=\emptyset$ 으로 설정.

Step 1: $i=i+1$ 로 설정하고 i 와 동일 또는 유사한 노드가 공급자 제품 카테고리에 존재하면 goto Step 2.

$i < n$ 이면서 i 와 동일 또는 유사한 노드가 존재하지 않으면 Step 1을 반복하고, $i=n$ 이면 알고리즘 종료.

Step 2: 매칭되는 공급자 제품 카테고리 노드를 m 이라 하고 이를 집합 P 에 추가.

만약 $i=1$ 이면 goto Step 3.

$1 < i < n$ 이면 goto Step 4.

$i=n$ 이면 goto Step 5.

Step 3: 노드 m 의 형제 노드 중에서 참조 노드와 동일 또는 유사한 노드가 존재하면 이를 집합 P 에 추가.

Step 4: $i=i+1$ 로 설정하고 노드 m 의 부모 노드 중에서 i 와 동일 또는 유사한 노드가 존재하면 goto Step 2.

$i < n$ 이면서 i 와 동일 또는 유사한 노드가 존재하지 않을 경우 Step 4를 반복하고, $i=n$ 이면 goto Step 5.

Step 5: 공급자 제품 카테고리에서 P 에 속한 노드들을 연결하는 sub-tree(sub category) 생성.

위의 매칭 알고리즘에서의 Step 0은 초기화 단계로서 인덱스 i 및 매칭 관계가 성립하는 노드를 저장하는데 사용되는 집합 P 를 초기화한다. Step 1에서는 구매자 카테고리라 처음으로 매칭되는 제품 카테고리 노드를 찾고, 3가지 경우로 나누어 각 경우에 해당하는 절차를 수행한다. 즉, 만약에 $i=1$ 인 검색 노드와 일치하는 노드가 존재하는 경우 Step 3으로 가서 목표 노드의 형제 노드를 대상으로 매칭되는 노드를 찾는다. 목표 노드와 매칭되는 노드가 존재하지 않을 경우 Step 3의 과정은 생략되고, Step 4를 통해 매칭 노드를 찾는 과정을 반복한다.

<그림 6>은 매칭 알고리즘을 사용하여 제품 카테고리 3을 분리한 결과를 나타낸다. 그림에서 보듯이 카테고리 3에 속하는 노드 중에서 구매자 카테고리의 검색 노드와 일치 또는 유사한 노드는 ‘memory’와 ‘electronic device’ 밖에 존재하지 않는다. 또한 목표 노드 ‘memory’와 매칭되는 형제 노드는 ‘lens’ 밖에 없으며, 이는 참조 노드인 ‘CPU’와 매칭되지 않음을 알 수 있다. 따라서 ‘memory’와 ‘electronic device’를 연결하는 sub-tree를 구성하면 <그림 5>의 우측과 같은 분리된 제품 서브 카테고리가 완성된다.

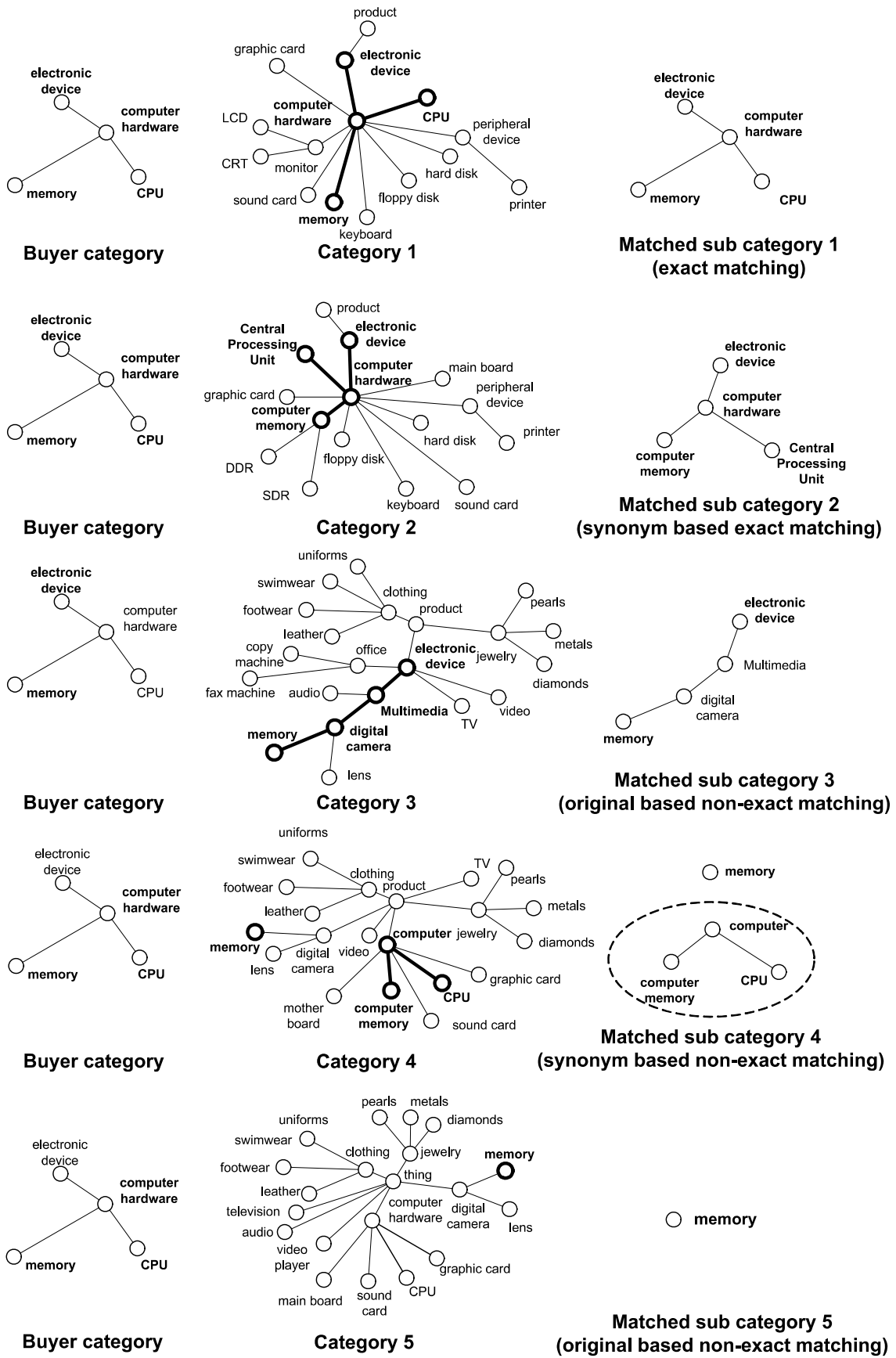


그림 5. 구매자 카테고리 와 공급자 제품 카테고리의 매칭 예제.

방식으로 구매자 카테고리 자신의 언어적 가중치를 구하기 위해 $(n+1) \times (n+1)$ 의 정방행 매칭 행렬을 구성하여 노드 i 의 언어적 가중치를 구하면 모든 경우에 대해서 $i=1$ 에서부터 n 인 노드는 1, $i=0$ 인 노드는 0.5의 언어적 가중치를 갖게 된다. 따라서 구매자 카테고리 자신의 언어적 척도 값은 $n+0.5$ 가 됨을 알 수 있다.

본 논문에서 제안하는 정규화된 언어적 척도 값 LM은 공급자 서버 카테고리의 구성하는 노드의 언어적 가중치 합과 구매자 카테고리 구성하는 노드의 언어적 가중치의 합의 비율로써 수식 (2)와 같이 정의되며, 이 값이 1에 가까울수록 공급자 제품 카테고리가 구매자 카테고리 와 언어적으로 유사하다고 볼 수 있다.

$$LM = \frac{\sum_j w_j^l}{1.0 \times n + 0.5} \quad (2)$$

<그림 4>의 구매자 카테고리인 경우, ‘electronic device’와 ‘computer hardware’, ‘memory’를 검색 노드로, ‘CPU’를 참조 노드로 결정했을 때, 각 검색 노드는 수식 (1)의 가중치 분류에 의해 SameAs이므로 가중치 1이 부여되고, 참조 노드 역시 SameAs이지만 수식 (1)에 따라 가중치 0.5가 부여된다. 따라서 구매자 카테고리 자신의 언어적 척도 값은 3.5가 된다.

<그림 5>의 공급자 서버 카테고리 3의 경우, ‘electronic device’와 ‘memory’는 “Search & SameAs” 관계이므로 가중치 값이 1이며, 나머지 노드는 “Nothing” 관계이므로 언어적 가중치 값이 0이 된다. 그러므로 수식 (2)에 의해서 제품 카테고리 3을 제공하는 공급자의 경우 $LM=2/3.5=0.571$ 이 된다. <그림 5>의 공급자 제품 카테고리 2의 경우, ‘electronic device’와 ‘computer hardware’는 “Search & SameAs” 관계이므로 가중치 값이 1이며, 목표 노드 ‘memory’와 참조 노드 ‘CPU’는 WordNet에서 정의한 ‘Central Processing Unit’과 ‘computer memory’의 유의어이므로 각각 “Search & Synonym” 및 “Reference & Synonym” 관계이다. 따라서 매칭된 서버 카테고리의 언어적 가중치 값은 3.2이고, $LM=3.2/3.5=0.914$ 이다. 정리하면, <그림 5>의 공급자 서버 카테고리 2가 구매자 카테고리 와 정확히 일치하진 않았지만 의미적으로 유사한 용어를 사용하였으므로 <그림 5>의 공급자 서버 카테고리 3보다 1에 근접한 값을 얻을 수 있었다.

4.3.2 구조적 척도(structural measure)

구조적 척도란 구매자 카테고리 와 공급자 서버 카테고리를 매칭할 때, 구매자 카테고리의 트리 구조와 공급자 서버 카테고리의 트리 구조간의 유사도를 측정할 것을 의미한다. 공급자 서버 카테고리가 주어졌을 때, 본 논문에서는 다음과 같은 방식으로 공급자 서버 카테고리의 노드 j 에 대한 구조적 가중치를 정의한다.

$$w_j^s = \begin{cases} 5, & \text{if } j = 0 \\ 10, & \text{if } j = 1 \\ w_{j-1}^s - \frac{w_{j-1}^s}{n}, & \text{otherwise} \end{cases} \quad (3)$$

만약 참조 노드와 관련된 $j=0$ 인 노드가 존재하면 이 노드의 구조적 가중치는 $w_0^s = 5$ 로 고정하여 설정한다. 검색 노드와 관련된 노드 1에서 m 까지에 대한 구조적 가중치는 $w_1^s = 10$ (목표 노드의 가중치)에서부터 j 를 증가시켜가면서 재귀적으로 구한다. 수식 (3)에서 목표 노드와 관련된 노드의 구조적 가중치는 최하위 노드에서 가장 높고, 상위 노드로 갈수록 그 값이 줄어들음을 알 수 있다. 이는 구매자가 찾고자 하는 제품이 하위 노드의 의미와 보다 밀접하게 연관되어 있기 때문에, 하위 노드로 갈수록 구조적 가중치를 크게 설정한 것이다. 사실 수식 (3)에서 가중치의 절대 값은 큰 의미가 없고, 참조노드와 목표노드 가중치 값 사이의 상대적 차이가 중요하다.

모든 공급자 서버 카테고리 노드의 구조적 가중치를 구조적 척도 계산시 반영하는 것보다는 구매자 카테고리 와 상관관계가 존재하는 노드들만을 대상으로 하여 구조적 척도를 계산하는 것이 보다 효과적인 것이다. 따라서 공급자 서버 카테고리의 노드 j 와 구매자 카테고리에 대한 상관관계의 존재 여부를 나타내는 지시 변수 (indicator variable)을 y_j 라 정의하면 y_j 는 다음과 같이 표현할 수 있다.

$$y_j = \begin{cases} 0, & \text{if } w_j^l = 0 \\ 1, & \text{if } w_j^l > 0 \end{cases} \quad (4)$$

본 논문에서는 이와 같은 지시 변수를 이용하여 공급자 서버 카테고리에 대한 구조적 가중치의 합을 $\sum_j w_j^s \cdot y_j$ 로 나타낸다. 마찬가지로 수식 (3)을 이용하여 구매자 카테고리의 노드 i 에 대한 구조적 가중치 uw_i^s 를 계산할 수 있으며, 공급자 서버 카테고리 와 달리 구조적 가중치의 합을 계산할 때 지시 변수를 고려하지 않아도 된다. 따라서 구매자 카테고리 자신의 구조적 가중치의 합은 $\sum_i uw_i^s$ 이 된다.

본 논문에서 제안하는 정규화된 구조적 척도 값 SM은 언어적 척도 값과 마찬가지로 방식으로 공급자 서버 카테고리를 구성하는 노드의 구조적 가중치의 합과 구매자 카테고리를 구성하는 노드의 구조적 가중치의 합의 비율로써 정의하며, 이 값이 1에 가까울수록 공급자 제품 카테고리가 구매자 카테고리 와 구조적으로 유사하다고 볼 수 있다.

$$SM = \frac{\sum_j w_j^s \cdot y_j}{\sum_i uw_i^s} \quad (8)$$

<그림 7>은 구매자 카테고리의 구조적 척도 값을 구하기 위해 각각의 노드에 구조적 가중치와 <그림 5>에 있는 공급자 서버 카테고리 2와 3의 각 노드에 부여된 구조적 가중치 값을 나타낸 것이다. 위의 값은 수식 (3)에 의해서 구할 수 있다. 따라서 서버 카테고리 2와 3의 정규화된 구조적 척도 값은 각각 $SM_1=189.202/189.202=1$, $SM_2=113.764/189.202=0.601$ 이 된

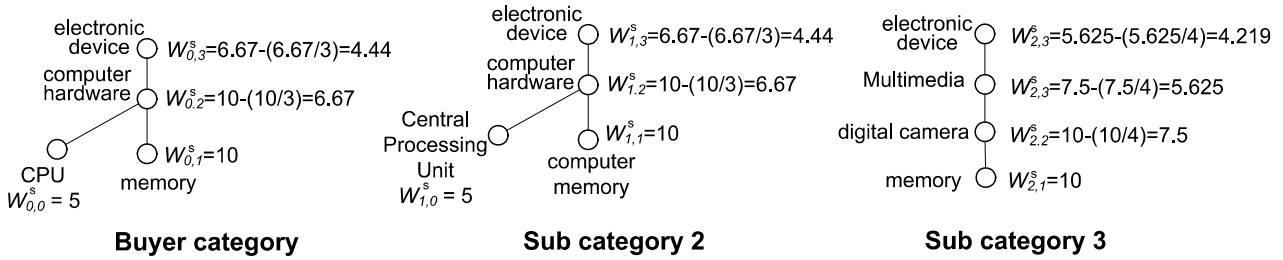


그림 7. 구매자 카테고리의 구조적 가중치($w_{0,j}^s$)와 공급자 서브 카테고리 2와 3의 구조적 가중치($w_{1,j}^s, w_{2,j}^s$).

다. 앞에서 살펴본 예제를 정리하면 <그림 5>의 공급자 서브 카테고리 2의 정규화된 언어적 척도 값은 0.914를, 정규화된 구조적 척도 값은 1을 얻었고, 공급자 서브 카테고리 3의 정규화된 언어적 척도 값은 0.571 정규화된 구조적 척도 값은 0.601을 얻었다. 의미적 연관성 척도 값은 두 척도 값의 합으로 정의된다. 언어적 척도 값과 구조적 척도 값은 0 이상 1 이하이므로 의미적 연관성 척도 값은 2에 근접할수록 구매자 카테고리 2와 유사하다는 것을 의미한다. 공급자 서브 카테고리 2와 3의 의미적 연관성 척도 값은 각각 1.914, 1.172이므로 공급자 서브 카테고리 2가 구매자 질의에 더 부합된다는 것을 알 수 있다.

5. Empirical Results

5.1 Implementation

이번 장에서는 본 논문에서 제안하는 구매자 카테고리 기반 지능형 전자상거래 메타 검색 엔진(Buyer category-based intelligent e-commerce meta-search engine)의 구현에 대하여 살펴보겠다. 본 논문의 실험을 위해 온톨로지 표현 언어 중 하나인 OWL을 사용하여 5개의 공급자 전자 카탈로그를 표현하였고, 카탈로그를 Java API와 JENA API를 이용하여 트리 구조로 변경하였다. GUI(Graphic User Interface)는 JavaSwing을 사용하여 구현하였다.

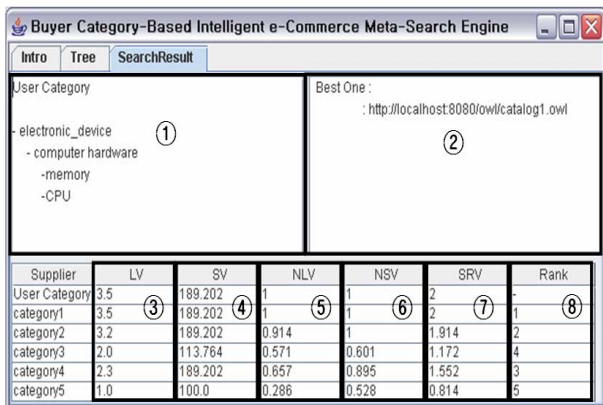


그림 8. 결과 화면.

<그림 8>은 구매자 카테고리(①)와 각각의 공급자가 제공하는 제품 카테고리의 비정규화된 언어적 척도 값 (③), 비정규화된 구조적 척도 값 (④), 정규화된 언어적 척도 값 (⑤), 정규화된 구조적 척도 값 (⑥), 의미적 연관성 척도 값 (⑦)과 의미적 연관성 척도 값에 따른 공급자 추천 순위(⑧)를 보여주고 있다. 또한 구매자 카테고리 질의 의도에 가장 부합한 전자 카탈로그를 가진 공급자의 웹 사이트 주소(②)를 구매자에게 보여준다.

5.2 Experimental Results

본 장에서는 제안하는 메타 검색 시스템의 타당성을 평가하기 위해 다음과 같은 두 개의 검색 시스템을 추가적으로 정의하였다.

카테고리 기반 의미 검색(category based semantic search) 확장 검색 노드 삽입 기능을 제거한 시스템으로, 구매자의 카테고리를 WordNet을 이용하여 공급자 제품 카테고리 2와 매칭하는 시스템이다.

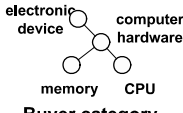
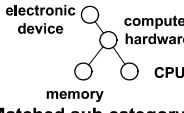
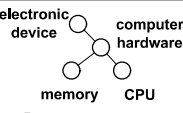
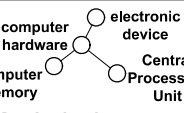
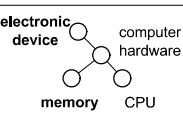
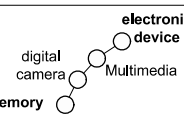
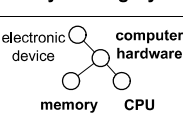
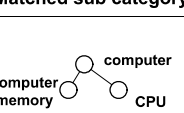
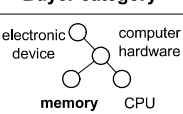
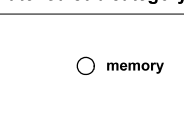
일반 키워드 검색(keyword based search) 구매자 카테고리에 있는 용어들을 단순히 AND 조합으로 생각하여 공급자 제품 카테고리 2와 매칭하는 시스템으로 공급자 제품 카테고리 2와의 구조적 연관성은 생각하지 않고 용어들의 일치만 고려하여 공급자 서브 카테고리 2를 분리하는 시스템이다.

<그림 5>에서 정의한 다섯 개의 공급자 제품 카테고리에 대해서 제안된 방법론을 적용한 검색 결과와 카테고리 기반 의미 검색과 일반 키워드 검색 시스템을 적용한 검색 결과를 비교한다. <표 2>는 제안된 검색, 카테고리 기반 의미 검색, 키워드 검색을 적용했을 때 구매자 카테고리 2와 다섯 개의 공급자 전자 카테고리의 유사성을 평가한 표이다.

사실, <그림 5>에서 알 수 있듯이 구매자의 의도가 컴퓨터 부품인 'memory'라면 제품 카테고리 1, 2, 4를 가진 공급자를 추천해야 좋은 검색 엔진이라 할 수 있다. 제품 카테고리 3과 5는 컴퓨터 메모리를 공급하지 않는 공급자이다. <표 2>에서 보듯이 제안된 검색 방법은 제품 카테고리 3과 5를 제공하는 공급자들의 추천 순위를 낮게 한다. 그리고 서브 카테고리 1은 구매자 카테고리 2와 공급자 서브 카테고리의 exact matching으로 모든 검색 방법에서 가장 높은 순위를 보이고 있다.

제안된 검색 방법과 비교해 볼 때 일반 키워드 검색은 용어의 유사도와 용어의 구조를 고려하지 않고 평가하기 때문에 매우

표 2. 검색 결과 비교

Search Method	Buyer category & supplier sub category	Linguistic Point	Structural Point	Semantic Relevancy	Rank	
Proposed search method	 Buyer category	 Matched sub category 1	1	1	2	1
	 Buyer category	 Matched sub category 2	0.914	1	1.914	2
	 Buyer category	 Matched sub category 3	0.571	0.601	1.172	4
	 Buyer category	 Matched sub category 4	0.657	0.895	1.552	3
	 Buyer category	 Matched sub category 5	0.286	0.528	0.814	5
Keyword based search	Buyer query: memory & electronic device & CPU Matched nodes: memory, electronic device, CPU	-	-	-	1	
	Buyer query: memory & electronic device & CPU Matched nodes: electronic device	-	-	-	3	
	Buyer query: memory & electronic device & CPU Matched nodes: memory, electronic device	-	-	-	2	
	Buyer query: memory & electronic device & CPU Matched nodes: CPU	-	-	-	3	
	Buyer query: memory & electronic device & CPU Matched nodes: memory, CPU	-	-	-	2	

나쁜 검색 결과를 보여준다. 일반 키워드 검색의 경우, 언어적 척도와 구조적 척도가 의미 없기 때문에 우선순위는 매칭된 용어의 수 순으로 정하였다. 그 결과, 서브 카테고리 3과 5의 우선순위가 두 번째로 높음을 알 수 있다. 이런 나쁜 결과는 용어간의 구조적 관계를 고려하지 않았기 때문에 발생한다.

정리하면, <표 2>에서 보듯이 일반 키워드 검색은 구매자 카테고리의 용어를 통해서만 검색하므로 검색 정확성이 가장 낮다. 서브 카테고리 3의 경우 구매자가 의도한 computer hardware의 memory와 상관없는 digital camera의 memory를 서브 카테고리 4의 결과인 computer memory보다 높게 평가하는 오류가 존재한다. 반면 제안된 검색 방법은 구매자 카테고리의

의미를 명확하게 해주기 위하여 시스템에서 자동으로 구매자 카테고리를 확장하므로 카테고리 기반 의미 검색 보다 높은 정확성을 보여준다.

6. 결론

인터넷의 발전과 더불어 온라인상에서 다양한 형태의 상거래를 수행하는 전자상거래의 중요성이 부각되고 있다. 하지만 현재의 전자상거래 시스템은 방대한 제품 정보 중에서 구매자가 원하는 제품을 정확히 검색할 수 있는 기능이 미흡한 실정

이다. 본 연구에서는 이러한 전자상거래의 문제점을 해결하기 위하여 구매자가 전자상거래 사이트에서 제품 카탈로그를 보다 효율적으로 검색할 수 있게 해주는 구매자 카테고리 기반 지능형 전자상거래 메타 검색 엔진(Buyer category-based intelligent e-commerce meta-search engine)을 제안하였다. 하지만 본 연구에서 제안하는 시스템은 용어의 확장을 전적으로 WordNet에 의지하고 있기 때문에 WordNet에서 정의한 용어와 용어간의 관계만을 따르게 된다. 따라서 제품 카탈로그에 특화된 용어들 간의 상위어, 하위어, 동의어를 정의하는 표준화 연구가 지속적으로 이루어져야 할 것이다. 그리고, 구매자가 카테고리를 편리하게 만들 수 있는 구매자 카테고리 생성 지원 도구에 대한 연구도 진행되어야 한다. 마지막으로 메타 검색 엔진이 제품 카테고리를 가지고 공급자를 추천하는 기능에서 더 나아가 각 공급자의 전자 상거래에 직접 연결되어 제품의 속성 정보도 함께 구매자에게 보여주는 시스템으로 확장되어야 할 것이다.

참고문헌

A. E. Howe, D. Dreilinger(1997), Savvy Search: A Metasearch Engine that Learns which Search Engines to Query, *AI Magazine*, 18(2), 19-25.
 B. Chandrasekaran and J. R. Josephson and V. R. Benjamins(1999), What Are Ontologies and Why Do We Need Them, *IEEE Intelligent Systems*, 14(1), 20-265. Copernic, Available at <http://www.copernic.com>.
 DCMI(The Dublin Core Metadata Initiative), Available at <http://dublincore.org>.
 Deborah L. McGuinness, Frank van Harmelen(2004), OWL Web Ontology Language Overview, W3C Recommendation 2004, Available at <http://www.w3.org/TR/owl-ref/>.
 Dogpile, Available at <http://www.dogpile.com/>.
 E-Commerce and Development Report(2003), UNITED NATIONS Conference on Trade and Development, Available at http://r0.uncad.org/ecommerce/docs/edr03_en/ecdr03.htm.
 E. Selberg, O. Ertzoni(1997), The MetaCrawler Architecture for Resource

Aggregation on the Web. *IEEE Expert*, 12(1) 11-14.
 F. Liu, C. Yu, W. Meng(2002), "Personalized Web Search by Mapping User Queries to Categories", *Conference on Information and knowledge management*, 558-565.
 G. A. Miller(1995), WordNet a Lexical Database for English, *Communications of the ACM*, 38(11), 39-41.
 H. Lee, J. Shim(2005), Conceptual and Formal Ontology Model of e-Catalogs, *Lecture Notes in Computer Science*, 3590, 11-20.
 J. Shim, Simon S. Y. Shim(2006), Ontology-based e-Catalog in e-commerce, *Electronic commerce research and applications*, 5(1), 1.
 Jena 2 A Semantic Web Framework, Available at <http://www.hpl.hp.com/semweb/jena.htm>.
 M. D. Mulvenna, S. S. Anand, A. G. Buchner(2000), Personalization on the Net using Web Mining, *Communications of the ACM*, 43(8), 123-125.
 M. Eirinaki, C. Lampos, S. Paulakis, M. Vazirgiannis(2004), Web Personalization Integrating Content Semantics and Navigational Patterns, *Web Information and Data Management*, 72-79.
 M. Eirinaki, M. Vazirgiannis, I. Varlamis(2003), SEWEP: Using Site Semantics and a Taxonomy to Enhance the Web Personalization Process, *Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining*, 99-108.
 M. Uschold, M. Gruninger(1996), Ontologies: Principles, Methods and Applications, *Knowledge Engineering Review*, 11(2), 93-155.
 N. Guarino, C. Masolo, G. Vetere(1999), OntoSeek: Content-Based Access to the Web, *IEEE Intelligent Systems*, 14(3), 70-80.
 S. Castano, A. Ferrara, S. Montanelli(2003), H-MATCH: an Algorithm for Dynamically Matching Ontologies in Peer-based Systems, *Semantic Web and Databases*, 231-250.
 S. Castano, A. Ferrara, S. Montanelli, D. Zucchelli(2003), HELIOS: a General Framework for Ontology-based Knowledge Sharing and Evolution in P2P Systems, *Database and Expert Systems Applications*, 597-6.
 S. Lawrence, C. L. Giles(1998), Context and Page Analysis for Improved Web Search, *IEEE Internet Computing*, 2(4), 38-46.
 UNSPSC(The United Nations Standard Products and Services Code), Available at <http://www.unspsc.org>.
 XTM(XML Topic Maps 1.0), Available at <http://www.topicmaps.org/xtm/1.0/>, August 2001.



김경필
 한성대학교 산업공학과 학사
 연세대학교 정보산업공학과 석사
 관심분야: SCM, BPM, 시멘틱 웹



김창욱
 고려대학교 산업공학과 학사
 고려대학교 산업공학과 석사
 Purdue University 산업공학과 박사
 현재: 연세대학교 정보산업공학과 교수
 관심분야: 정보시스템 설계, RFID 기반
 물류시스템, 웹 서비스



우상훈
 한성대학교 산업공학과 학사
 연세대학교 정보산업공학과 석사
 현재: i2 TECHNOLOGIES 재직 중
 관심분야: SCM, BPM, 시멘틱 웹, RFID 기반
 실시간 물류 정보시스템