분류와 속성 정보를 이용한 상품 검색 시각화 기법

Visualization Techniques for Product Searching : using classification and property information

강성희(Seonghee Kang)*, 심준호(Junho Shim)**

초 록

정보의 시각화는 정보의 직관적 이해를 위해 중요한 역할을 하며, 다양한 응용영역에서 해당 환경의 특징에 맞추어 활용되고 있다. 본 논문에서는 전자상거래에서 상품을 효과적으로 검색하여 보여주기 위해서는 기존의 방식이 아닌 상품 정보의 특색에 맞는 검색 방식과 시각화 기법이 필요하다고 판단하여, 기존의 MapNet이나 클러스터맵등의 기술을 반영한 상품 검색 비주얼라이제이션 기법을 소개한다. 상품 카테고리 검색과 속성 검색의 분리를 통해 검색조건에 명확한 시맨틱 부여를 가능하게 하고, 사용자에게 검색결과를 평가하고 재검색을 수행할 수 있게 한다. 결과 평가는 원하는 결과를 얻지 못했을 경우 사용자가 질의어를 재작성하여 최적의 결과를 찾을 수 있게 한다.

ABSTRACT

Information visualization plays an important role to provide a conceptual comprehension of the information and should be adjusted and featured to reflect the characteristics of the domain to which it is applied. Product searching in e-Commerce is not an exception. In this paper, we present visualization techniques that are specified for effectively browsing the results of product searching. We considered visualization techniques including MapNet and Cluster Map and modified them to reflect the characteristics of the product domain. We consider two types of search queries: category searching and product property searching, and make it possible to include more detailed product semantics in their search criteria. We also provide a query refinement mechanism so that even users with rack of understanding the products may rewrite their queries for better results.

키워드 : 전자카탈로그, 전자상거래, 시각화

e-Catalog, e-Commerce, Visualization

* 숙명여자대학교 석사 학위과정시 수행되었으며, 저자의 현 직장은 삼성전자 정보통신총괄 통신연구소임.

** 숙명여자대학교 컴퓨터과학과, 교신저자

1. 서 론

상품 정보의 공유성을 높이고 시스템에 의한 검색의 정확도를 높여 사용자에게 최적의 상품 검색 결과를 제공하기 위한 시도가 최근에 활발하다[1-4]. 그런데 자연어 입력의 키워드 검색이나 카테고리 선택식 검색이 대부분인 현재의 전자 카탈로그 검색 방식은 상품 정보의 특성을 제대로 반영한다고 보기 어렵다[1,5].

현재 대부분의 상품 검색은 자연어 검색을 채택하고 있다. 자연어 검색의 가장 큰 문제점은 사용자가 자연어를 사용해서 자신이 원하는 상품을 얼마나 잘 명세할 수 있는지에 있다. 즉, 사용자가 무엇을 사야겠다고 머릿속에 가지고 있다할지라도 그것을 제대로 명세하지 못하면 시스템이 원하는 상품을 찾아줄 확률은 적어지는 것이다.

[5]에서 제시한 사용자가 적절한 단어를 사용하지 못하는 경우를 상품 정보 검색에 적용해 살펴보면 문제가 되는 경우는 다음과 같다. 첫째, 원하는 상품의 적절한 명칭을 사용자가 전혀 떠올릴 수 없는 경우이다. 이 경우 적절한 단어를 찾기 위한 사용자의 여러 시행착오가 불가피하다. 둘째, 매우 일반적인 단어를 사용하여 검색 결과가 어쩔 수 없이 의미 없는 결과들을 많이 포함하는 경우이다. 마지막으로, 두 단어 이상의 조합에 논리 기호인 'AND'나 'OR'를 사용하는 경우이다. 논리적 조합에 익숙하지 않은 일반 사용자들에게 이것은 까다로운 문제가 될 수 있다.

기존 방식이 가지는 또 다른 문제점은 상품이 가지고 있는 다양한 속성에 대한 검색을 지원하지 않는다는 것이다. 여기서 상품이 가지는 속성이란 크기, 가격, 색상과 같이 상품이 가지는 특징을 말한다. 예를 들어, 조달청 상품 데이터베이스에서 '17인치 액정 모니터'를 검색하면 검색 결과로 130개의 상품이 리턴된다. 이보다 더 나아가 하나의 제조사의 속성에 '삼성'이라는 조건을 추가하여 검색하게 되면, '17인치 삼성 액정 모니터'로 54개의 상품을 검색 결과로 얻게 된다. 상품 속성의 검색 조건 하나의 추가로 반 이상의 상품들을 사용자의 고려대상에서 제외시킬 수 있게 된다. 그러나 속성 레벨의 검색이 아닌 상품명 위주의 검색을 지원하고 있는 많은 시스템에서 이 이상의 검색 조건을 지원하지 못하고 있는 실정이다. 설사 여러 속성 검색을 지원한다 할지라도, 현재 상품 데이터베이스에서는 상품마다 다른 속성명 또는 속성값, 단위(UOM : unit of measure)등을 사용하기 때문에 정확한 검색이 불가능할 수밖에 없다. 따라서 상품 검색 시스템에서는 이 다양한 속성을 검색에 적극 활용하여 좀 더 효과적인 결과를 도출할 수 있도록 하는 새로운 방식이 필요하다.

마지막으로 지적할 수 있는 문제점은 검색과 그 결과의 정제(refinement) 측면이다. 만일 사용자가 입력한 키워드 중 하나라도 만족되지 않으면 시스템은 결과가 없음을 나타내거나 몇 가지 조건만을 만족하는 결과를 보여주게되고, 이 경우 사용자는 자신이 작성한 질의의 어떤 문제로 원하는 결과를 찾을 수 없는지 알기 힘들다. 반면, 질의 결과가 너무 많은 경우도 문제다. 사용자가 원하는 결과를 찾기 위해 고려해야 할 결과가 많아지는 것은

최소의 클릭으로 원하는 결과를 찾게 하고자 하는 검색 시스템의 목적에서 벗어난다. 이처럼, 조건을 만족하는 결과가 아무것도 없거나, 반대로 사용자가 작성한 질의의 결과가 너무 많은 검색 결과를 나타낼 때 질의의 재작성이 필요한데, 두 경우 모두에 있어 아무 단서 없이 질의를 다시 작성하는 것은 사용자에게 부담이 될 수 있다. 즉, 검색 결과가 너무 많은 것은 특정 키워드 조건이 너무 많은 인스턴스를 포함한 경우일 수 있으며, 반면 검색 결과가 없는 것은 잘못된 용어 사용의 문제일 수도 있고, 조건이 미세하게 범위를 벗어났을 수도 있는데 기존의 시스템은 사용자에게 이 원인을 제시해주지 못하고 있다.

본 논문에서는 이러한 기존 검색 시스템의 문제점을 해결하고 상품 정보의 특성을 반영하여 더 나은 검색 결과를 보여줄 수 있는 상품 검색 방식을 제안하고자 한다. 제안할 검색 방식에서 고려된 사항은 다음과 같다 :

(1) 다양한 정보를 지닌 각종 상품을 제대로 찾기 위해 중요한 것은 검색하려는 상품의 정확한 명세이이며, 이를 위해서는 질의 검색어에 상품의 시맨틱 부여 메카니즘이 필요하다.

(2) 상품 데이타베이스는 통상적으로 하나의 상품 컨셉에 수많은 인스턴스를 포함한다. 상품 검색에서 중점적으로 보여줘야 할 대상으로서 상품의 분류를 위해 존재하는 컨셉보다는 개개의 상품인 인스턴스일 경우도 많다.

(3) 상품의 중요한 정보이자 각각의 구별을 가능하게 하는 다양한 속성에 대한 검색이 강화되어야 한다. 시스템이 다양한 속성을 검색 조건에 포함시킬 수 있어야 원하는 상품을 사용자가 아닌 시스템이 검색해낼 수 있다.

(4) 잘못 작성된 질의를 다시 작성할 수 있게 질의의 문제점을 사용자에게 알려줘야 한다. 더 나아가 올바른 속성값 선택을 통해 검색을 좀 더 간편히 할 수 있게 하는 질의 수정에 대한 추천(recommendation) 기법이 필요하다.

2. 연구 배경 및 관련 연구

2.1. 상품 정보와 전자카타로그

상품정보는 상품분류체계와 다양한 상품 속성 그리고 이들 간의 관계등의 풍부한 정보를 담고 있다. 비즈니스에 있어 거래 당사자 간에 의미적으로 잘 정리되고 풍부한 상품정보를 공유하고 교환하는 것은 매우 중요하며, 이러한 이유로 상품정보 도메인에 온톨로지를 적용하려는 시도는 e-비즈니스의 주요 연구영역으로 최근에 부각되어 왔다 [1,3-4,6-7]. [8]에서는 상품정보의 구축, 유지, 통합의 어려운 면을 설명하고 이에 대한 해결방안으로서 상품정보에 의미적 부여를 하는 방식을 제안하고 있다. [9]에서는 UNSPSC[1]와 eCl@ss[2]와 같은 표준산업 상품분류체계를 상품체계전반의 일반적이고 개념적인 상위

1) Http://www.unspsc.org

2) Http://www.eclass-online.com

(upper) 온톨로지로 사용하고 각 산업별로 세부적이고 구체적인 하위(lower) 온톨로지를 구축하는 방법론을 제시하였다.

의미적 모델링이 특정 도메인에 적합하게 구축되기 위한 선결조건은 해당 도메인에 존재하는 주요 컨셉과 컨셉 간의 관계를 명확히 분석해내는 것이다. 좀 더 기술적으로 말하자면 상품정보를 온톨지적으로 모델링한다는 것은 상품정보를 컨셉과 관계 그리고 이에 수반되는 제약사항(constraints)으로 명확히 개념화하는 시키는 것을 의미된다[8]. 본 논문에서 언급되는 의미적 관계들은 저자들이 부분적으로 참여한 [1]에서 분석 취득된 것에 기초하며, 수반된 개념화의 이론적 바탕이 되는 형식적 표현(formal representation)은 DL(description logics)을 사용하여 [10]에서 기술되었다. 본 논문에서 사용하는 상품정보의 시맨틱은 상품 온톨로지에서 존재하는 다양한 관계들을 모두 적용한 것은 아니며, 분류체계관계와 상품정보의 각종 속성의 활용에 그 초점을 두고 있다.

[6,4,1]는 기업현장에서 필요한 대규모 상품 온톨로지 데이터베이스 구축에 있어 OWL등을 사용한 지식베이스 방법론은 현실적으로 견고성(robustness)과 확장성(scalibility)에 문제가 있음을 지적하고, 온톨로지를 관계형 데이타베이스 방식으로 모델링하고 구축 운용하는 것이 기업 현장에서 바람직한 적용 방법론임을 주장하였다. [1,10]에서는 이에 따른 구체적인 모델링 기법을 제시하고 있다.

2.2. 비주얼라이제이션

비주얼라이제이션 기법들은 각각의 특징이 있는데 최적의 데이터 비주얼라이제이션을 위해서는 시각화하려는 데이터의 성격을 잘 파악한 후 그것에 맞는 비주얼라이제이션 기법을 선택하는 것이 중요하다. 본 연구자는 상품 정보 검색 환경에 적합한 일련의 비주얼라이제이션 기법을 개발하여 이를 추후에 궁극적으로 상품 온톨로지 환경으로 제공하려는 시스템 PROVIS (PRoduct Ontology VIsualization System)을 개발해 오고 있다. 본 절에서는 PROVIS에서 채택한 주요 비주얼라이제이션 기법을 중심으로 각 기법들의 특징을 살펴보고 개괄적으로 설명한다.

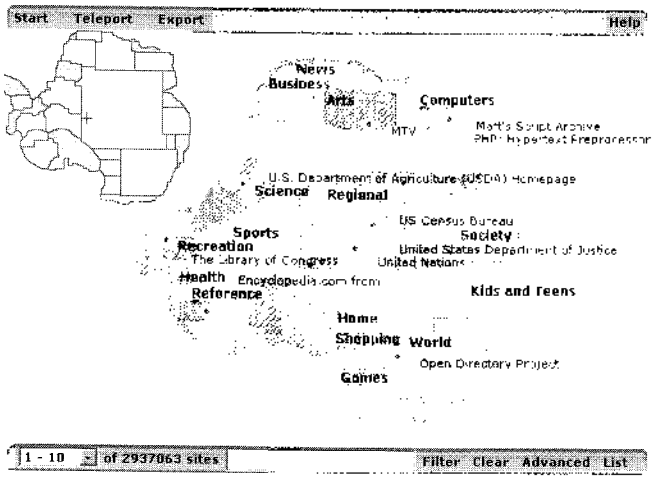
2.2.1. MapNet

Antarcti.ca's Visual Net[3] 에 기반해 ODP(Open Directory Project)[4]에 적용된 것이 <그림 1>에 보여지고 있는 MapNet이다. ODP는 카테고리를 만들고 인터넷 문서를 적절한 카테고리에 분류하여 시맨틱을 부여하기 위한 프로젝트로 그 분류 과정이 인간에 의해 이루어지고 있다. <그림 1>에서 네모 모양의 도형으로 표시된 것은 ODP에서 분류한 카테고리이고 타겟 모양은 인스턴스, 즉 그 카테고리를 대표하는 사이트가 된다. 즉 Arts 카테고리에 대표적인 인스턴스인 사이트로 MTV가 존재하는 것이다.

지도를 응용한 맵넷 기법은 트리 구조의 상위, 즉 일반적(general) 컨셉부터 보여주며 클

3) Antarctica System Inc., http://www.antarctica.net/

4) Open Directory Project, http://dmoz.org/, (directory.mozilla.org)라고도 불림



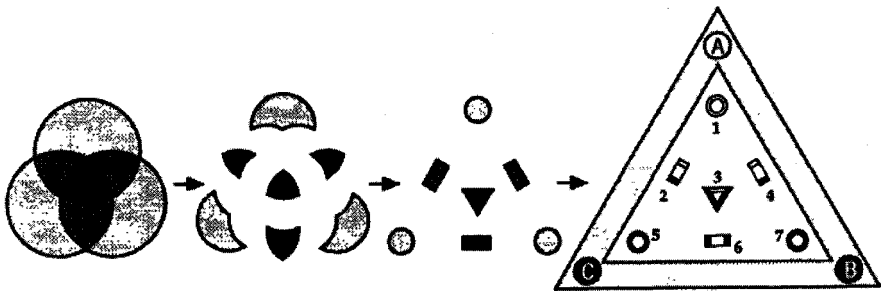
〈그림 1〉 맵넷의 예

릭 시 드릴다운(drill-down)되어 클릭된 컨셉의 서브 컨셉들, 즉 세부적(specific) 컨셉으로 점차 접근할 수 있게 한다. 그러므로 관심 있는 부분만 확대하여 볼 수 있는 장점이 있다. 이 기법은 특히 대용량 데이터를 표시하는데 좋으며, 화면 가운데 포커스를 두고 가장자리로 갈수록 데이터가 작게 보이는 Hyperbolic Tree[12] 기법과 비교될 수 있다. 또한 이 기법은 많은 컨셉과 인스턴스를 보여줄 수 있지만 둘 이상의 컨셉에 속한 인스턴스의 겹쳐짐은 표현하지 못한다는 것이 단점이다.

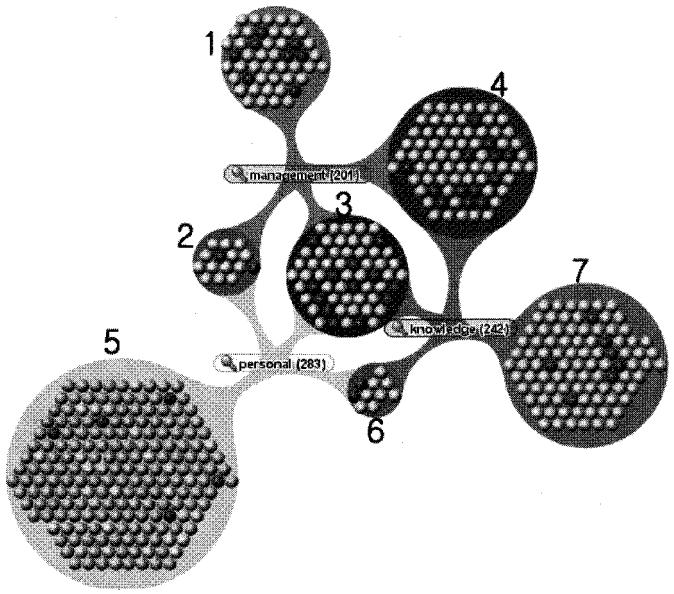
2.2.2. 클러스터 맵(Cluster Map)

[13]에서는 시맨틱 웹의 온톨로지의 특징으로 인스턴스의 수가 클래스의 수에 비해 훨씬 많다는 것을 언급했다. 클래스의 수가 수천 개 정도라면 인스턴스의 수는 수백만 개에 이를 수 있다는 것이다. 상품 온톨로지도 마찬가지이다. 같은 카테고리에 속한 상품이 카테고리 수에 비해 훨씬 많을 뿐만 아니라, 같은 속성을 가지는 상품의 수 또한 속성에 비해 매우 많다. 이것은 컨셉과 인스턴스를 한 화면에 표현할 때 있어 고려되어야 할 사항이다. 또한 상품 온톨로지 비주얼라이제이션에서 중요한 것은 컨셉이 아니라 상품인 인스턴스이다. 그러므로 인스턴스 중심의 비주얼라이제이션이 필요하다. 이 인스턴스 레벨의 비주얼라이제이션에서 고려해야할 중요한 특징은 여러 컨셉에 동시에 속하는 인스턴스를 어떻게 표현해야 할 것인가 이다. 이 컨셉 간의 겹쳐짐, 즉 조합을 잘 표현할 수 있는 것이 벤다이어그램(Venn diagram) 기법이다. 이 벤다이어그램의 단점은 3개 이상의 컨셉의 겹쳐짐을 표현하면 가시성이 많이 떨어진다는 점과 겹쳐지는 부분의 면적이 상당히 작아진다는 점이다.

이 벤다이어그램 기법을 잘 변형한 것이 InfoCrystal[14]이다. InfoCrystal은 3개 이상의 컨셉도 표현할 수 있다. 〈그림 2〉에서 바깥쪽의 A, B, C라 쓰여진 원 세 개는 각각 A, B.



〈그림 2〉 Infocrystal의 예



〈그림 3〉 클러스터 맵의 예

C 라는 컨셉을 나타낸다. 그리고 각각의 내부 아이콘의 모양은 원은 하나의 컨셉, 직사각형은 두 개의 컨셉, 삼각형은 세 개의 컨셉, 정사각형은 네 개의 컨셉에 겹쳐짐을 나타낸다. 그리고 1은 A and (not(B or C)), 2는 A and C and (not B), 3은 (A and B and C), 4는 A and B and (not C), 5는 C and (not(A or B)), 6은 B and C and (not A), 7은 B and (not(A or C))의 조합을 나타낸다고 볼 수 있다.

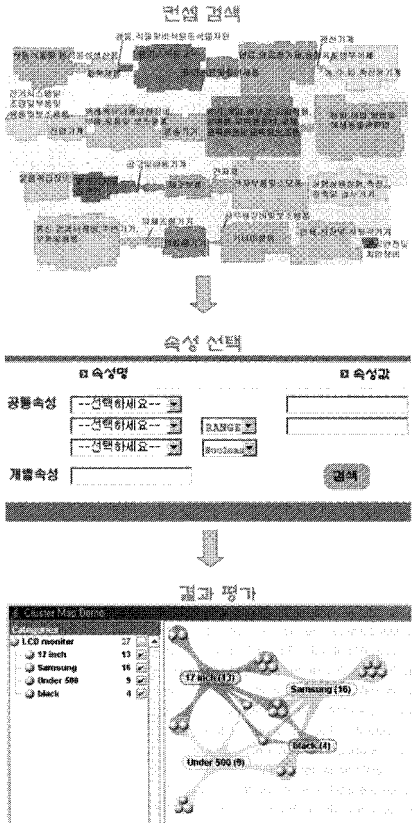
이 아이콘에 숫자를 표시함으로써 각각에 속한 인스턴스의 개수를 반영할 수는 있으나 개개의 인스턴스를 표현하는데 있어서는 이 기법도 역시 미흡하다. 작은 아이콘 안에 각각의 인스턴스를 표현하기는 어렵기 때문이

다. 이처럼 현재의 비주얼라이제이션 기법 중 인스턴스 레벨의 비주얼라이제이션이 가능하면서 많은 인스턴스들을 보여줄 수 있는 것은 거의 없다. 특히 3개 이상의 컨셉에서의 인스턴스의 겹쳐짐 표현에서 가시성이 현저히 떨어진다. 그러나 여러 컨셉의 겹쳐짐에서 인스턴스 레벨의 비주얼라이제이션을 가능케 하는 것이 클러스터 맵 기법이다.

〈그림 3〉에서 보여지는 클러스터 맵은 management, personal, knowledge의 3개의 클래스를 나타내고, 각 클래스에 연결된 풍선 모양의 클러스터들은 그 클래스가 포함하는 인스턴스들을 보여준다. 여기서 인스턴스는 클래스가 토픽(topic)인 문서가 된다. 그리고 클러스터 내의 다른 색의 인스턴스는 문서의 타입을 나타낸다. 즉, 웹페이지인지, 워드 파일인지에 따라 다른 색으로 표현되는 것이다. 또한 두 개 이상의 클래스에 연결된 클러스터는 두 개 이상의 클래스에 동시에 포함되는 인스턴스들을 나타낸다. 이와 같은 방식으로 클러스터 맵은 인스턴스 간의 조합(combination)을 보여준다[15].

InfoCrystal과 마찬가지로 클래스의 조합을 분석해 보면 1은 management and (not(knowledge or personal)), 2는 management and personal and (not knowledge), 3은 (management and knowledge and personal), 4는 management and knowledge and (not personal), 5는 personal and (not(management or knowledge)), 6은 knowledge and personal and (not management), 7은 knowledge and (not(management or personal))의 조합을 나타낸다고 볼 수 있다.

3. PROVIS 시스템

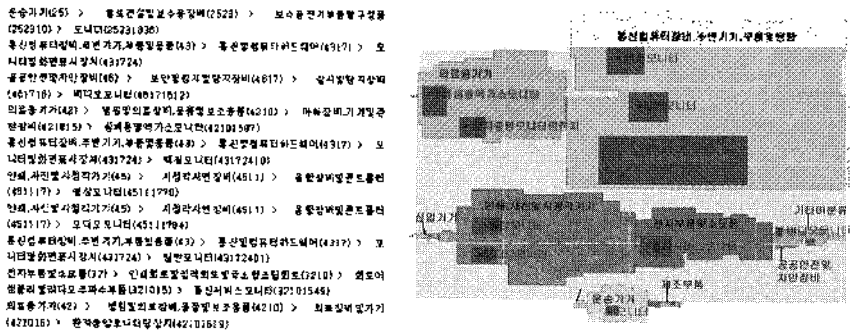


〈그림 4〉 PROVIS 시스템 전개도

〈그림 4〉는 상품 정보 검색 시각화를 위해 개발 중인 PROVIS 시스템 전개도이다. 시스템은 컨셉 검색, 속성 선택, 결과 평가의 세 개의 단계로 이루어져 있으며 다음에서 차례로 각 단계에 대해 살펴볼 것이다.

3.1. 컨셉 검색

일반 상품 검색 시스템이 보통 키워드 검색



〈그림 5〉 G2B 분류에서 '모니터'로분류를 검색한 결과(좌)와
맵넷에서 '모니터'로 상품 컨셉을 검색한 결과(우)의 비교

에서 컨셉 전체, 즉 상품(products), 분류체계(classification scheme), 속성(attributes)과 단위체계(UOMs)를 포괄 대상으로 하는 반면, 본 시스템에서는 상품이 속한 카테고리를 나타내는 분류체계(classification scheme) 컨셉만을 검색하는 질의를 따로 분리하여 처음 단계에서 검색하게 한다. 이는 일반 키워드 검색에서 컨셉과 속성을 구분하지 않고 한꺼번에 입력받음으로 인해 얻게 되는 쓸모없는 검색 결과들을 줄이기 위한 것이다. 이 분류체계 컨셉의 검색에 2.3.1절에서 소개한 맵넷 기법을 사용했는데 이는 서론에서 시스템 고려사항 첫 번째로 언급한 사항 때문이다. 즉, 맵넷을 통해 검색어에 시맨틱을 부여해 주고자 하는 것이다. 시스템의 입력 데이터로 사용한 조달청 온톨로지는 분류체계로 주로 G2B분류를 사용한다. 이 G2B분류는 컨셉이 총 4개의 레벨로 이루어져 있는 트리 구조이다. 이 컨셉 트리의 리프 노드인 깊이가 4인 레벨의 노드에 인스턴스들이 속하게 된다.

〈그림 5〉의 좌측 그림은 검색어로 '모니터'를 입력한 경우 G2B 분류에 대한 검색 결과를 트리 형식의 텍스트로 표현한 것이다. 이것으로 알 수 있듯, G2B 상품분류체계의 검색에서 '모니터'로 컨셉을 검색하게 되면, 모니터는 컴퓨터의 디스플레이 장치뿐만 아니라 철도장비, 감시및탐지장비, 의료용기기, 음향장비및컨트롤러, 전자부품및소모품등 여러 카테고리에서 발견된다. 이것은 서론에서 언급한 부적절한 검색어의 사용에 대표적인 예라고 볼 수 있다. 사용자는 자신이 사려고 하는 것에 대한 개념화가 있음에도 불구하고 '모니터'라는 단어와 같이 너무 일반적인 명세를 한 것이다. 이는 조달청 상품 데이터베이스 같이 다양한 카테고리의 방대한 양의 데이터를 다루는 경우 더 두드러지는 문제가 될 수 있다.

〈그림 5〉의 좌측 결과를 맵넷으로 표현한 것이 우측 그림이다. 시각적으로 훨씬 보기 편해진 것을 알 수 있다. 그러나 시각적 효과를 떠나 맵넷을 사용하는 보다 중요한 이유는 사용자의 부정확한 카테고리 명세를 명확히 하기 위한 것이다. 즉, 화면에 모니터와 일치하는 카테고리들을 보여줌으로써 사용자의

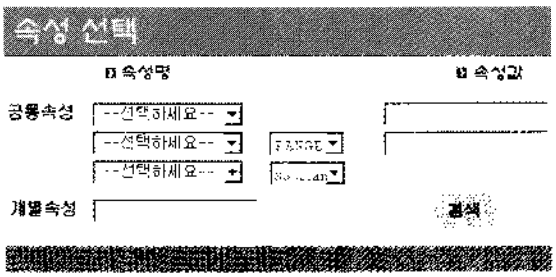
클릭을 통해 검색어의 보다 명확한 명세를 유도하는 것이다. 예로 들면, 맵넷에서 결과로 보여준 '모니터' 컨셉이 속한 카테고리 중 하나를 클릭함으로써 어느 모니터인지를 지정하게 되는데, 이때 사용자가 액정 모니터를 선택하게 되면 〈통신컴퓨터장비〉〈주변기기〉〈부품및용품〉〈통신및컴퓨터하드웨어〉〈모니터및화면표시장치〉〈액정모니터〉모니터〈/...〉와 같이 검색어에 시맨틱을 부여하게 되는 것이다. 이렇게 되면 사용자의 검색어는 기계가 그 의미를 명확하게 이해할 수 있는 형태로 명세되게 된다. 이런 명확한 명세는 기계가 부정확하게 명세된 검색어에 대해 광범위하게 그 의미를 추론하여 결과적으로는 사용자가 원치 않았던 결과까지 포함하게 되는 경우를 제거하기 위해 반드시 필요한 과정이다. 즉, '모니터' 라는 사용자의 검색어에 대해 시스템이 컴퓨터의 디스플레이 장치뿐만 아니라 철도용 모니터, 감시용 모니터, 의료용 모니터, 음향장비 모니터 등 여러 카테고리를 검색한 결과를 보여준 후, 사용자에게 원하는 것을 선택하게 하는 것이 아니라, 시스템이 사용자가 원하는 것을 명확하게 인식하고 검색 범위를 한정 시킨 후 검색을 수행해야 검색 효율을 높일 수 있을 뿐만 아니라, 검색 결과의 정확성도 높일 수 있게 되는 것이다.

3.2. 속성 선택

컨셉 선택을 통해 인스턴스를 어느 정도 줄일 수 있었다고 할지라도 아직 고려해야할 인스턴스가 많다. 본 시스템에서 원하는 인스턴스를 선택하는데 고려되는 요소들은 속성이다. 즉, 속성으로 컨셉 내 인스턴스들을 구별할 수 있다.

〈그림 6〉은 사용자로부터 원하는 상품의 속성명과 그 속성의 속성값을 입력받기 위한 인터페이스이다. 이 때 속성은 크게 공통속성과 개별속성으로 나눠지는데 공통속성은 선택된 상품 컨셉 내의 인스턴스들이 공통으로 가지는 속성이고, 개별속성은 공통속성 외에 특정 상품만이 가지는 속성이다.

공통속성의 속성명은 콤보박스에 나타나고 사용자는 그 중에서 고려하고자 하는 속성명을 선택하고 속성값도 입력한다. 공통속성은 다시 속성값이 특정 값을 가지는 것과 범위를 가지는 값으로 나눠지는데, 범위를 가지는 속성은 상품의 가격이나 크기 같은 것으로 보통



〈그림 6〉 속성 선택 창

특정 상수 값으로 입력하기 어려운 것들이다. 이 때 범위(range)로 사용되는 것이 이상, 이하 등이고 범위는 선택사항이다. 만약 가격이 40만원에서 50만원 사이의 상품을 원한다면 속성명은 가격, 속성값은 40만원, 범위는 이상을 우선 택하고, 다음으로 속성명 가격, 속성값 50만원, 범위 이하를 택하면 되기 때문에 하나의 속성명에 굳이 두 개의 범위와 범위값을 두진 않았다. 이 외에도 속성값이 Boolean인 경우는 속성값을 콤보박스로 TRUE, FALSE로 선택하게 한다.

마지막으로 개별속성의 경우는 특정 상품에만 존재하는 속성명을 가지는 것으로, 위 공통속성의 속성명에 원하는 속성명이 없을 경우 텍스트 입력으로 검색이 가능하게 한다. 예를 들어, mp3 플레이어 기능이 가능한 디지털카메라를 검색하고자 한다면 mp3 플레이어가 디지털카메라란 컨셉의 공통속성으로 존재하지 않고 특정 상품에만 국한된 속성이기 때문에 개별속성으로 검색하게 된다.

3.3. 결과 평가

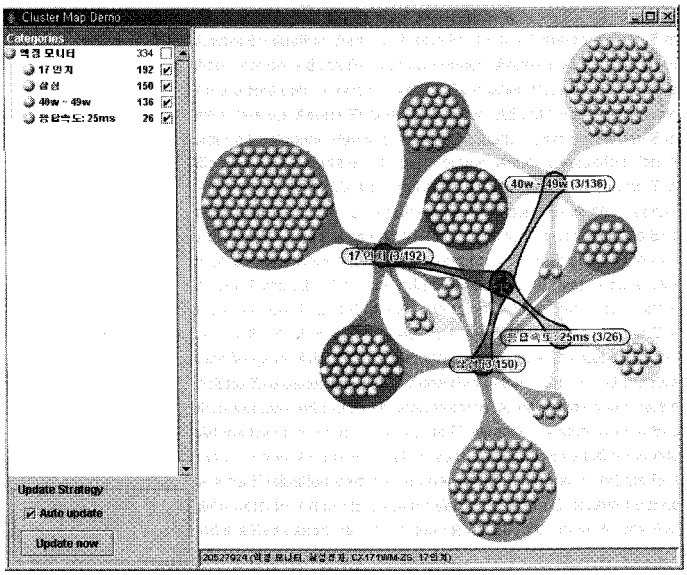
사용자가 속성을 선택하면 그 결과는 클러스터 맵(Cluster Map)으로 보여진다. 클러스터 맵은 사용자가 상품 컨셉 검색에서 선택한 컨셉의 인스턴스들 중 해당 속성을 만족하는 인스턴스들을 각 속성당 클러스터로 나타낸다. 즉, 컨셉과 인스턴스를 한 화면에 나타내면서 각 인스턴스까지 세세히 화면에 보여줄 수 있는 기법이 클러스터 맵인 것이다.

클러스터 맵은 한 화면에 최대 7개의 클러스터를 표현할 수 있는데, 사용자가 7개 이상의 속성을 선택할 가능성이 적을 뿐만 아니라, 컨셉 검색 단계를 거치면서 화면에 나타날 인스턴스의 수가 많이 줄어들기 때문에 가독성 역시 떨어지지 않는다. 이는 여러 클러스터에 중복되어 있는 인스턴스라도 한 화면에 한번 밖에 나타나지 않기 때문이다.

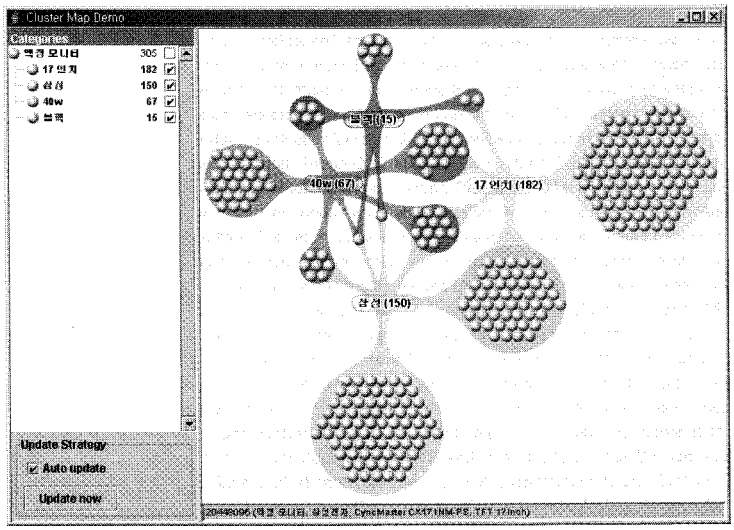
〈그림 7〉은 클러스터 맵으로, 컨셉 검색에서 액정 모니터를 선택한 후 속성 검색에서 화면크기 17인치, 제조사 삼성, 소비전력 40W대(40W~49W), 응답속도 25ms를 입력하여 이를 만족하는 액정 모니터 컨셉 내의 인스턴스들을 속성별 클러스터로 묶어서 보여준 것이다.

클러스터 맵은 사용자가 어렵게 생각할 수 있는 AND나 OR 연산 같은 기호를 질의문에 사용하지 않고도 그림을 통해 그 관계를 유추할 수 있게 한다. 예를 들어, A AND B는 A와 B에 동시에 연결된 클러스터를 보면 된다. 〈그림 7〉을 통해 사용자는 입력한 4가지 속성에 동시에 연결된 클러스터를 찾아 조건을 모두 만족하는 인스턴스를 검색할 수 있게 된다. 〈그림 7〉에서는 모든 조건을 만족하는 3개의 상품이 검색된 것을 볼 수 있으며, 각 인스턴스에 마우스를 가져다 대면 화면 하단의 바에 상품 정보가 나타나게 된다.

결과 평과 단계에서 클러스터 맵을 사용한 이유는 입력한 속성값을 모두 만족하는 결과가 없거나 그 결과가 만족스럽지 않을 경우 클러스터 맵을 사용하면 질의의 정제(Query refinement)가 가능하기 때문이다. 즉, 질의 후 결과(result set)가 공집합이 나오거나 너무 클 수가 있다. 이 경우, 클러스터 맵으로 나타난 결과를 보면 제시된 조건을 모두 만족하는 결과뿐만 아니라, 부분적으로 만족하는 결과까



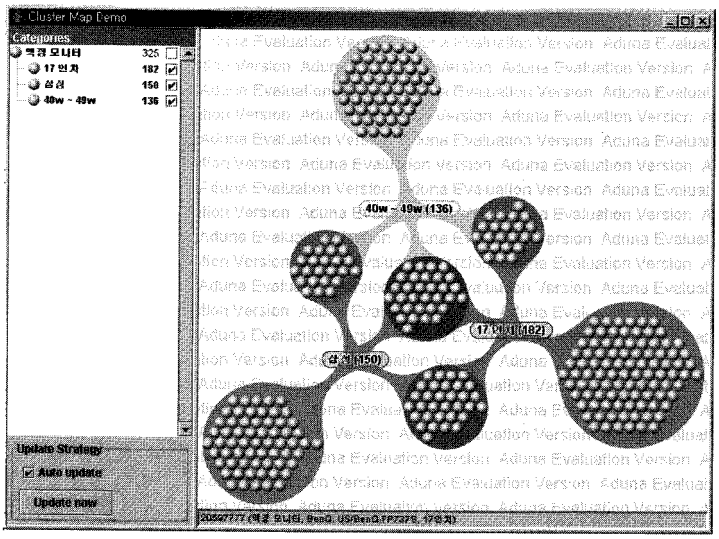
〈그림 7〉 클러스터 맵을 이용한 결과 평가



〈그림 8〉 클러스터 맵 - 결과가 없는 경우

지 한눈에 볼 수 있기 때문에 질의를 수정하는 것이 용이해진다.

〈그림 8〉는 결과가 공집합이 나온 경우다. 이는 특정 속성을 만족하는 인스턴스들이 사용자가 원하는 속성의 일부를 포함하지 못하는 경우로 질의 조정 시 질의를 좀 더 일반화



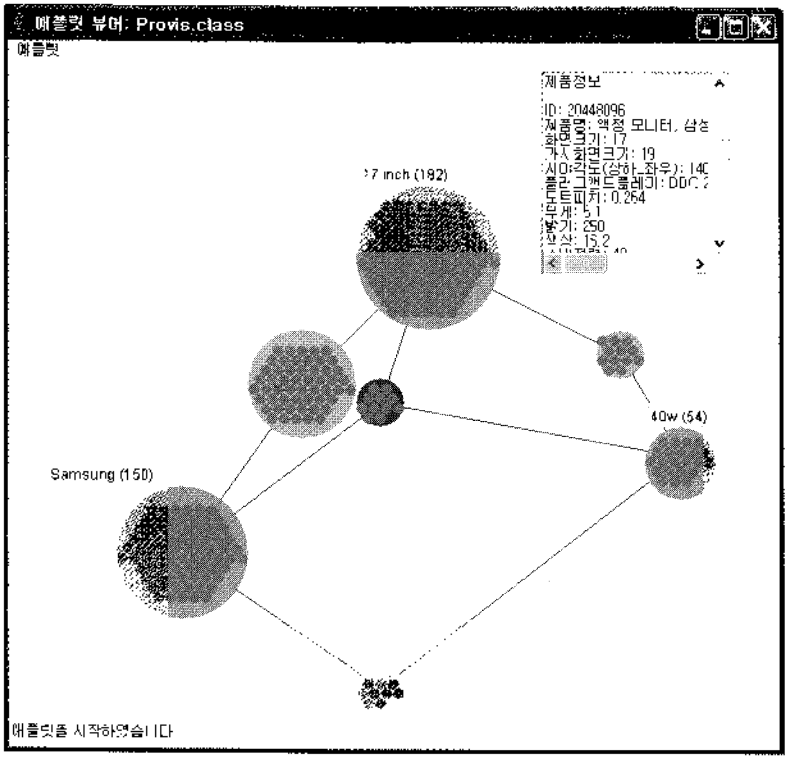
〈그림 9〉 클러스터 맵 - 결과가 너무 많은 경우

해야 한다. 그러기 위해서는 특정 속성을 그 속성을 포함하는 부모 속성으로 대체하는 것이 일반적이다. 그러나 일반 키워드 검색에서는 어떤 속성을 대체해야 하는 것인지를 결정하기가 쉽지 않다. 어떤 속성에 대한 검색이 어느 정도의 인스턴스를 결과로 리턴하는지 알 수 있는 방법이 없기 때문이다.

〈그림 8〉의 예를 살펴보자. 사용자가 화면 크기 17인치, 제조사 삼성, 소비전력 40W, 색상 블랙을 모두 만족하는 상품을 찾고자 하기 때문에 4개의 속성에 모두 연결된 클러스터가 있어야 하나, 결과가 존재하지 않는다. 이 경우, 사용자는 클러스터 맵 결과화면을 보고 결과 중 최대 개수인 3개의 속성을 동시에 만족하는 클러스터들을 참고하여 나머지 속성을 조정함으로써, 수정된 조건을 포함한 새로운 질의를 구성하면 이를 만족하는 인스턴스를 찾아낼 수 있을 것이란 것을 알게 된다. 이 예에서는 40W, 17인치, 삼성의 조건을 만족하는 클러스터를 통해, 만족되는 인스턴스를 얻기 위해 색상 속성의 속성값인 블랙을 조정할 필요가 있음을 알 수 있게 된다.

〈그림 9〉처럼 결과가 너무 많은 경우도 문제다. 이것은 액정 모니터 중 크기 17인치, 제조사 삼성, 소비전력 40W ~ 49W인 것을 찾는데 만족된 인스턴스가 많은 경우로 이를 일일이 다 클릭해서 결과를 찾는 것 대신 질의의 구체화를 통해 결과를 좀더 걸러낼 수 있다. 그러기 위해서 몇몇 클래스를 좀더 구체화된 서브 클래스로 대체할 필요가 있다. 이 예제의 경우는 결과 화면을 통해 40W ~ 49W를 40W로 바꾸거나 하는 등의 대안이 가능함을 알 수 있다.

추천(recommendation)은 해당 클러스터가 나타내는 속성을 그대로 유지한 상태에서 속성값의 변경에 대한 제안을 제시하는 것이다. 이때, 제안될 속성값은 해당 속성값이 수치일 경우와 스트링일 경우로 나뉘어 고려되어야



〈그림 10〉 오차를 이용한 결과를 포함한 클러스터 맵

한다. 50인치 TV를 찾는 예를 들어보자. 이는 속성값이 수치인 경우인데, 이때 수치 approximation이 추천이 될 수 있을 것이라 생각할 수 있다. 예를 들어, 50인치 TV가 존재하지 않는다고 하면 시스템에서는 우선 50인치에 가까운 크기를 가지는 TV를 제안할 수 있을 것이다. 그러므로 수치에 $\mp\alpha$를 한 속성값을 바탕으로 재검색을 제안하는 것이 가능하다. 다음은 수치값이 스트링인 경우로 빨간색 액정 모니터를 찾는 예를 생각해 볼 수 있다. 이때 빨간색인 액정 모니터가 거의 없거나, 빨간색이면서 사용자가 제시한 나머지 조건들을 만족하는 액정모니터가 거의 없으면, 추천으로 가능한 것은 빨간색을 포함하는 더 큰 속성 컨셉 또는 흰색, 검정처럼 다수의 액정 모니터가 가지는 색이 될 것이다.

다음으로 본 시스템에서 고려한 것은 제시된 속성값에 대해 미세한 오차를 허용하는 것이다. 이것은 주로 수치값에 대한 오차가 될 것이다. 만일 사용자가 원하는 가격 속성의 속성값을 5만원이라고 입력했다고 했다면 기존 시스템에서 50,000원과 정확히 일치하는 가격의 상품만 검색 결과로 보여준다면, 본 시스템에서는 50,100원이나 49,900원까지도 검색 결과에 포함하여야 한다는 것을 반영하였다. 물론 어느 범위까지 오차를 허용할 것인

가는 앞으로 고려되어야할 사항일 것이다.

〈그림 10〉은 오차를 반영한 클러스터 맵을 결과로 보여준 것이다. 이때 소비전력 40W를 사용자의 입력값으로 받았으나, 41W의 액정 모니터까지도 결과에 반영할 경우, 제시된 값과 정확히 일치하는 상품이 진분홍색으로 표현된 반면, 오차를 허용함으로써 포함된 부분에 대해서는 녹색으로 나타내었다. 이는 시스템이 제시한 오차를 허용하고 싶지 않으면 진분홍색 인스턴스만을 고려대상으로 이용할 수 있게 하기 위해서이다.

4. 논의 사항

클러스터 맵을 개발한 Aduna에서는 Aduna Cluster Map Library를 제공하기 때문에 이를 이용해 클러스터 맵 평가 버전을 이용할 수 있다. 본 연구에서는 그 평가버전에 상품 정보 데이타를 입력 데이터로 사용하여 결과를 테스트 하였다. 그러나 평가 버전이 기본적인 클러스터 맵 이상의 기능을 제공하지 못해 상품 온톨로지의 특성을 반영한 보다 효율적 검색을 위하여 부분적인 수정이 불가피하여 이를 위해 클러스터 맵을 JAVA 언어로 다시 구현하였다. 현재 PROVIS 시스템은 이 클러스터 맵 부분만이 구현된 상태이기 때문에, 이전 단계인 컨셉 검색과 속성 선택 단계는 이미 수행되었다는 가정 하에 입력 데이터를 프로그램에 직접 연동해 구현하였다.

현재 구축된 PROVIS 시스템의 경우 조달청 상품정보 온톨로지 데이터베이스의 일부를 입력 데이터로 사용하였으며, 이것은 관계형 데이터베이스로 구축되어 있다. 입력 데이터는 관계형 데이터베이스로부터 SQL 질의를 통해 원하는 데이터를 검색하는 형태를 취한다. 향후에는 상품의 의미적 정보가 대표적 온톨로지 언어인 RDF, OWL 언어로 표현된 온톨로지로 된 경우를 입력으로 처리하도록 확장시킬 예정이다.

이와 관련한 추가적인 이슈는 시각적 모델링 도구와의 연계이다. 상품 온톨로지 시각화는 상품 온톨로지 모델링, 온톨로지 데이터베이스의 구축등과 연계되어질 때 그 효과가 극대화될 수 있다. 본 연구자는 이를 위한 관련 연구로서 Protege[5]나 I*Com[6]과 같은 시각적 에디팅 기능을 가지는 상품 온톨로지 모델링 툴[16] 을 개발중이며, 본 논문에서 제안하는 시스템과 통합하려는 계획을 가지고 있다.

5. 결 론

본 논문에서는 전자상거래에서 상품 검색과 검색결과를 효과적으로 시각화하기 위한 고려 요소를 언급하고 이에 대한 해결 기법을 제시하였다. 맵넷을 활용한 컨셉 검색은 사용자가 다양한 상품 정보 데이터베이스에서 원하는 상품 컨셉을 명확히 명세할 수 있게 하였으며, 속성 검색을 통해 특정 속성관계를 가지는 인스턴스를 찾아내는 일련의 과정을 지원해준다. 다양한 속성을 통한 검색은 온톨

5) http://protege.stanford.edu
6) http://www.inf.unibz.it/~franconi/icom/

로지 개념을 이용하여 속성명의 상이함으로 인해 배제되었던 검색 결과까지 포함할 수 있게 되어, 기존 시스템보다 더 나은 결과를 기대할 수 있게 한다. 마지막으로 클러스터 맵 기법은 사용자가 화면에 나타난 결과를 분석하고 이를 바탕으로 질의를 수정할 수 있도록 하는 '추천'을 제공하여 상품에 대한 해박한 지식 없이도 질의의 수정을 가능케 한다.

향후 과제로 고려되어야 할 부분으로서는 상품 온톨로지의 장점을 좀 더 활용할 수 있는 것에 대한 관련 이슈들이다. 예를 들어, 온톨로지가 컨셉과 컨셉 간의 관계를 나타내는 것이라는 점에서 컨셉 간의 관계를 어떻게 그래픽적으로 표현할 것인가가 중요한 이슈가 될 수 있다. 현재 시스템에서는 맵넷 상에서 선으로 컨셉들을 연결하여 컨셉 간의 관계를 나타내고 있으나, 너무 많은 관계가 존재할 때 가시성이 상당히 떨어질 수 있음이 문제가 될 수 있을 것이다. 따라서 사용자가 어떤 관계를 검색하고 싶은지를 구체적으로 명세하고, 시스템이 명세된 관계만을 화면에 표시해 주는 것이 효율적일 것이다. 예를 들어, '책상과 공동구매 관계인 컨셉'과 같이 구체적인 질의에 대해 책상 컨셉과 공동구매 관계인 의자 컨셉을 서로 특정 색상의 선으로 연결하여 공동구매 관계를 나타낼 수 있다. 그리고 이런 관계의 표현은 더 나아가 인스턴스 간의 관계 표현에도 활용될 수 있으며 이것은 클러스터 맵 위에서 표현 가능할 것이다.

또한 상품 온톨로지 분야 연구에서 구체화되고 있는 동의어, 이의어 관계, 보완품, 대체품, 부속품과 같은 다양한 관계를 검색에 활용하고 이들에 대한 검색 결과를 효과적으로 시각화하는 통합 기법에 대한 연구 확장이 필요하다.

참 고 문 헌

[1] I.-h. Lee, S. Lee, T. Lee, S.-g. Lee, D. Kim, J. Chun, H. Lee, and J. Shim, "Practical Issues for Building a Product Ontology System", International Workshop on Data Engineering Issues in E-Commerce (DEEC2005), IEEE Computer Society, 2005.

[2] W. Kim, D. W. Choi, and S.-u. Park, "Intelligent Product Information Search Framework Based on the Semantic Web", 3rd International Semantic Web Conference (ISWC2004), Springer-Verlag, 2004.

[3] J. Leukel, V. Schmitz, and F. Dorloff, "A Modeling Approach for Product Classification Systems", 13th International Conference on Database and Expert Systems Applications (DEXA 2002), Springer-Verlag, 2002.

[4] S. Lee, "Design & Implementation of an e-Catalog Management System", Tutorial at 9th International Conference on Database Systems for Advances Applications (DASFAA2004), 2004.

[5] B. Shneiderman, "The Eyes Have It : a Task by Data Type Taxonomy for

Information Visualizations", 1996 IEEE Symposium on Visual Languages (VL' 96), IEEE Computer Society, 1996.

[6] J. Lee and R. Goodwin, "Ontology Management for Large-Scale E-Commerce Applications", Electronic Commerce Research and Applications, Elsevier, 2006.

[7] M. Hepp, "Measuring the Quality of Descriptive Languages for Products and Services", in E-Business - Standardisierung und Integration, F.-D. Dorloff, et al(Editors), Tagungsband zur Multikonferenz Wirtschaftsinformatik, 2004.

[8] D. Fensel, Y. Ding, B. Omelayenko, E. Schulten, G. Botquin, M. Brown, and A. Flet, "Product Data Integration in B2B E-Commerce", IEEE Intelligent Systems, Vol. 16, No. 4, IEEE Society, 2001.

[9] L. Obrst, R.E. Wray, and H. Liu, "Ontological Engineering for B2B E-Commerce", International Conference on Formal Ontology in Information Systems (FOIS'01), ACM Press, 2001.

[10] H. Lee, J. Shim, and D. Kim, "Ontological Modeling of e-Catalogs using EER and Description Logic", International Workshop on Data Engineering Issues in E-Commerce (DEEC2005), IEEE Society, 2005.

[11] 조달청 온톨로지 상품검색 시스템, http://www.g2b.go.kr:8100/index.jsp

[12] J. Lamping, R. Rao, and P. Pirolli, "A Focus+context Technique based on Hyperbolic Geometry for Visualising Large Hierarchies", ACM Conference on Human Factors in Software, ACM Press/Addison-Wesley Publishing Co., 1995.

[13] C. Fluit, M. Sabou, and F. van Harmelen, "Supporting User Tasks Through Visualisation of Light-weight Ontologies", Handbook on Ontologies, Springer, 2004.

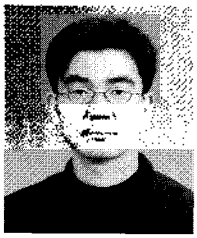
[14] A. Spoerri, "InfoCrystal : a Visual Tool for Information Retrieval and Management", 2nd International Conference on Information and Knowledge Management, ACM Press, 1993.

[15] Cluster Map, http://aduna.biz/index.html.

[16] K. Kim, M. Tark, H. Lee, J. Shim, J. Lee, and S. Lee, "PROMOD : A Modeling Tool for Product Ontology", 2nd International Workshop on Data Engineering Issues in E-Commerce and Services (DEECS2006), Springer-Verlag, 2006.

저 자 소 개

강성희 (E-mail : diane@sookmyung.ac.kr)
2003. 2. 숙명여자대학교 정보과학부 컴퓨터과학전공 학사
2006. 2. 숙명여자대학교 컴퓨터과학과 이학석사
2006. 2 ~ 현재 삼성전자 정보통신총괄 통신연구소 차세대단말팀 선행기술 Lab
관심분야 데이타베이스, 온톨로지



심준호 (E-mail : jshim@sookmyung.ac.kr)
1990. 2. 서울대학교 계산통계학과 학사
1994. 2. 서울대학교 계산통계학과 전산과학전공 이학석사
1998. 12. Northwestern Univ. USA, Electrical & Computer Eng. 공학박사
1999. 2 ~ 1999. 11. Computer Associates Int'l. USA, R&D Staff
1999. 11 ~ 2001. 2. Drexel Univ. USA, Assistant Professor
2001. 3 ~ 현재 숙명여자대학교 정보과학부 컴퓨터과학전공 부교수
관심분야 데이타베이스, 전자상거래, 상품정보, 온톨로지