

# 전자상거래 환경에서의 분류체계 자동 통합 기법

김재범<sup>o</sup> 김동규 이상구  
서울대학교 컴퓨터공학부  
{retiger, dkkim, sglee}@europa.snu.ac.kr

## Automatic Category Merging Technique in Electronic Commerce

Jaebom Kim<sup>o</sup> Dongkyu Kim Sanggoo Lee  
School of Computer Science and Engineering, Seoul National University

### 요 약

인터넷은 단순한 정보 교환의 수단이 아니라 기업들의 이윤 창출을 위한 새로운 공간이 되고 있으며 수많은 쇼핑몰들이 이를 설명해 주고 있다. 하지만 분류체계 측면에서 각 쇼핑몰들이 제공하고 있는 분류체계에는 크게 다음 두 가지의 문제점이 있다. 첫째로 각 쇼핑몰마다 서로 다른 자기만의 상품 분류체계를 가지고 있다는 점이다. 이로 인해 쇼핑몰을 이용하고자 하는 사용자는 각 쇼핑몰을 방문할 때 마다 혼란스러울 수 밖에 없다. 두 번째는 각 쇼핑몰이 제공하고 있는 분류체계는 정적인 형태만을 띄고 있다는 점이다. 따라서 사용자는 이미 정해져 있는 상품에 대한 분류의 체계만을 좋건 싫건 간에 따라야 한다. 따라서 이러한 문제들을 해결하기 위하여 본 논문에서는 규칙이라는 추가 정보를 가지도록 모델링된 쇼핑몰의 분류체계들에 대하여 자동적인 통합의 기법을 제시한다. 제시된 기법에 의하여 쇼핑몰 사용자들에게 모든 쇼핑몰의 통합된 뷰의 제공, 사용자별 분류체계의 생성, 메타 쇼핑몰 간의 통일된 인터페이스 제공 등을 할 수 있다.

### 1. 서론

인터넷이 우리 생활 깊숙이 파고든 지금, 인터넷은 단순한 정보 교환의 수단이 아니라 기업들의 이윤 창출을 위한 새로운 공간이 되고 있다. 이러한 현상은 하루에도 여러 개씩 생겨나고 있는 인터넷 쇼핑몰들을 보면 쉽게 알 수 있다. 인터넷 쇼핑몰을 굳이 두 종류로 나누어 보면 일반 쇼핑몰과 여러 쇼핑몰의 상품을 한꺼번에 검색할 수 있게 해주는 메타 쇼핑몰로 나누어 볼 수 있다. 이러한 쇼핑몰들은 자신들이 취급하고 있는 상품에 대한 효율적인 검색을 위하여 상품에 대한 분류체계를 제공하고 있다는 공통점이 있다.

분류체계의 통합이라는 측면에서 일반 쇼핑몰이 아닌 메타 쇼핑몰의 분류체계에 대하여 초점을 맞추어 보자. 현재 서비스 중인 메타 쇼핑몰 분류체계의 한계는 다음과 같다. 첫째, 메타 쇼핑몰은 여러 일반 쇼핑몰의 상품 검색이 가능하게 하기 위하여 각 쇼핑몰의 분류체계를 통합한 거대한 분류체계를 가지고 있다고 볼 수 있으며 이러한 통합 과정에서 많은 분류체계의 정보가 손실된다는 점이다. 즉, 예를 들어 A쇼핑몰과 B쇼핑몰의 모든 상품에 대한 검색이 가능한 메타 쇼핑몰이 있을 때 A쇼핑몰에 존재하는 특정 상품에 대한 분류체계 상의 탐색 경로가 메타 쇼핑몰의 분류체계에서는 존재하지 않는 경우가 많다는 것이다. 둘째, 메타 쇼핑몰을 이용하는 개별

사용자들은 서로 다른 관심 분야를 가짐에도 불구하고 각 사용자마다 서로 다른 분류체계에 대한 뷰를 제공하지 못하고 있다는 점이다. 셋째, 새로운 메타 쇼핑몰이 하나 생길 때 마다 나름의 방식에 의하여 나름의 분류체계를 만들어야 하는 중복된 노력이 필요하다는 점이다. 마지막으로 이러한 서로 다른 분류체계를 가진 메타 쇼핑몰들은 서로 간의 거래에서도 어려움을 가지게 된다는 점이다. 이러한 모든 문제들은 분류체계의 통합을 위한 완벽하고 통일되고 자동화된 기법이 없다는 데서 발생하는 것이다.

따라서 이러한 여러 문제들을 해결하기 위하여 쇼핑몰의 분류체계에 대한 자동 통합 방법은 꼭 필요하게 된다. 이러한 분류체계 자동 통합 기법에 의해 얻을 수 있는 장점은 다음과 같다. 첫째, 쇼핑몰 사용자들에게 모든 쇼핑몰의 통합되고 정형화된 뷰를 제공할 수 있다. 둘째, 자동 통합 기법을 사용자별 맞춤 분류체계 생성에 이용할 수 있다. 셋째, 분류체계 측면에서 쇼핑몰 간에 통일된 인터페이스를 제공할 수 있다. 넷째, 이러한 자동 통합 기법은 분류체계가 쓰이고 있는 모든 분야 즉, web community, 통합 신문 사이트, VOD & AOD 사이트 등에서 이용 가능하다는 것이다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구에 대한 소개를 하고, 3장에서는 자동 통합 알고리즘의 기본 바탕이 되는 분류체계의 모델링 방법과 분류체계 각 노드의 추가 정보인 규칙에 대해 설명한다. 이어서 4장에

\* 이 논문은 2000년도 두뇌한국21사업에 의하여 지원되었음.

서는 자동 통합 알고리즘에 대하여 설명하고, 5장에서는 이러한 자동 통합 알고리즘의 구현에 대한 설명이 있을 것이다. 끝으로 6장에서 결론과 향후 과제에 대하여 기술할 것이다.

2. 관련연구

인터넷 쇼핑물의 분류체계에 관한 연구는 SNUDCL 시스템에서 동적이고 유연하고 유지가 용이한 상품 분류체계 모델링 방법과 질의어를 제시한 바 있으며 이러한 분류체계 모델링 방법이 본 논문의 기본 바탕이 되고 있다[1][2].

그리고 지금 서비스 되고 있는 많은 인터넷 쇼핑물들을 보면 분류체계는 거의 모두 다음과 같은 특징을 가진다.

- 트리의 형태이며 트리의 각 노드는 별도의 추가 정보를 가지지 않는다.
- 트리 구조의 표현은 각 노드들 간의 물리적인 링크를 이용한다.
- 트리의 레벨은 4-5정도이다.
- 각 쇼핑물 마다 서로 다른 구조를 가진다.

이러한 형태를 띄고 있는 분류체계에서는 각 노드들 간의 부모-자식간의 관계를 파악할 논리적인 정보가 없기 때문에 자동적인 통합 기법을 고안할 수 없다.

그 외에 현재 나와 있는 여러 EC Solution들을 보아도 동적인 형태의 분류체계나 분류체계에 대한 자동 통합의 기능은 제공하고 있지 않다.

3. 분류체계 모델링과 분류체계 노드의 규칙

앞서 언급한 바와 같이 분류체계 통합 알고리즘은 SNUDCL 시스템의 분류체계 모델링 방식을 이용한다 [1][2]. SNUDCL 시스템에서는 분류체계가 그래프의 형태를 띄고 있으며 각 노드는 규칙이라는 추가의 정보를 지니게 된다. 여기서의 규칙은 각 노드에 대한 설명 또는 언어로서 분류체계 노드와 상품들을 연결시켜 주는 predicate의 역할을 하는 동시에 노드들의 부모-자식 관계를 논리적으로 알아내기 위한 핵심 정보가 된다. 규칙은 기본적인 상품군 정보와 추가적인 디멘전 정보에 의하여 구성된다. 여기서 상품군 정보는 가장 기본이 되는 상품의 그룹을 의미하는 것으로 예를 들면 ‘노트북’, ‘데스크탑PC’, ‘에어컨’, ‘전자레인지’와 같은 것이다. 이에 관한 규칙을 상품군 소규칙( $R_g$ )이라 하자. 그리고 디멘전 정보는 상품의 검색을 용이하게 할 목적으로 쓰이는 것으로 ‘사용자’, ‘가격’, ‘계절’, ‘브랜드’ 등이 해당된다. 즉 디멘전 정보에 의해서 하나 또는 그 이상의 상품군에 포함된 상품을 더 작은 집합으로 나눌 수 있는 것이다. 이에 관한 규칙을 디멘전 소규칙( $R_d$ )이라 하자. 이러한 요소로 규칙을 구성한 이유는 일반적으로 모든 상품은 특정 상품군에 속한다고 볼 수 있고, 다시 이러한 상품군의 상품들은 보고자 하는 관점에 따라서 작은 집합으로 나누어 질 수 있기 때문이다. 즉, 보고자 하는 관점이 디멘전에 해당하는 것이다.

그래서 이러한 규칙이라는 것을 고려하여 분류체계를

구성하는 과정을 보면 다음과 같다.(괄호안은 분류체계 노드 이름과 규칙)

- 기본적인 상품군을 결정한다.  
(예: TV<상품군=TV>, 오디오<상품군=오디오>)
- 상품군들의 유사성에 기초를 두어 상품군의 계층 구조를 생성한다.→ 상품군들의 or연결  
(예: 전자제품<상품군=TV or 상품군=오디오>)
- 보고자 하는 관점에 따라 상품군을 나눈다. → 상품군 집합과 디멘전 정보를 and로 연결  
(예:LG오디오<<상품군=오디오> and <브랜드=LG>>)

결국 이러한 규칙의 형태를 고려해 볼 때 분류체계 각 노드의 규칙은 다음과 같은 정규식 형태로 일반화 할 수 있다.

$$(R_g(\text{or } R_g)^*(\text{and } R_d))^*$$

즉, 분류체계의 한 노드의 규칙은 상품군에 관한 소규칙이 하나 또는 그 이상이 ‘or’로 연결되고 필요에 따라 하나 또는 그 이상의 디멘전에 관한 소규칙이 ‘and’로 연결될 수 있다는 것이다. 이러한 규칙 구성의 이유는 앞서 설명한 분류체계 구성 과정을 보면 알 수 있다.

4. 분류체계 자동 통합 알고리즘

분류체계 자동 통합 알고리즘은 3장에서 설명한 규칙이라는 추가 정보와 모든 쇼핑물의 분류체계는 이러한 규칙을 추가 정보로 가진 분류체계의 형태로 변환할 수 있다는 두 가지 점에서 출발한다. 이러한 기본 사항을 기초로 하여 분류체계 자동 통합의 단계를 살펴보면 다음 그림 1과 같다.

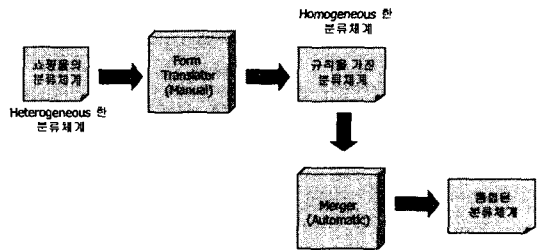


그림 1 분류체계 자동 통합의 과정

두 분류체계의 통합이라는 것은 결국 각각의 분류체계가 가지고 있던 노드들의 부모-자식 간의 관계를 그대로 유지시켜 주면서 하나의 거대한 분류체계로 만들어 주는 것이다. 이러한 과정에서 필수적으로 필요한 정보가 바로 각 노드들의 부모-자식관계를 판별하는데 필요한 정보이고 이를 위하여 앞서 설명한 노드의 규칙을 이용하게 된다. 규칙을 이용하게 되면 두 노드가 있을 때 각 노드가 가진 규칙의 포함관계를 따짐으로써 어느 노드가

부모이고 어느 노드가 자식에 해당하는 지를 쉽게 알 수 있는 것이다.

이러한 분류체계의 자동 통합을 수행하는 과정에서 반드시 지켜져야 하는 점은 첫째로는 정보의 손실이 없는 통합이 되어야 한다는 점이고, 둘째로는 통합되기 이전의 상태로 쉽게 복구 가능해야 한다는 점이다. 이 같은 사항은 통합 알고리즘이, 각각의 노드가 가진 규칙의 포함관계에 의한 계층 구조를 파괴하지 않으면서 통합을 시도하기 때문에 가능하게 된다.

통합 알고리즘은 크게 세 개의 단계로 구성이 된다. 첫번째 단계는 search space를 줄이기 위한 과정에 해당되고, 두 번째 단계는 각 노드의 통합될 최적 위치를 결정하는 과정에 해당되고, 마지막 세 번째 단계는 각 case별 노드의 통합 적용의 과정에 해당된다. 통합알고리즘의 첫번째와 두 번째 단계를 state diagram으로 표현하면 그림 2과 그림 3과 같다.

이 같거나 또는 한쪽이 다른 쪽에 포함되는 경우만 생기게 된다. 따라서 각 경우에 맞게 동일한 노드로 취급하거나 아니면 부모-자식의 관계를 고려하여 추가하거나 하게 된다. 그림 4는 이러한 분류체계 자동 통합 적용의 예이다.

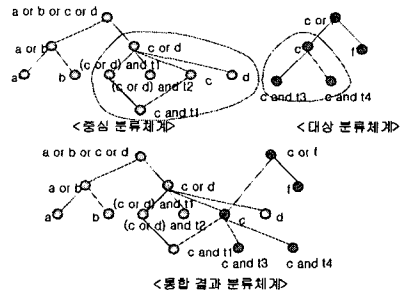


그림 4 분류체계 통합의 예

• State Diagram : Phase1

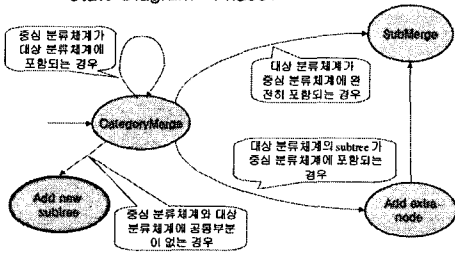


그림 2 통합 알고리즘 1단계

• State Diagram : Phase2

◆ 대상 분류체계의 각 노드들에 대하여 아래 작업 수행

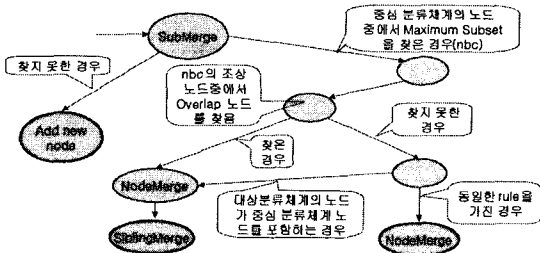


그림 3 통합 알고리즘 2단계

두 노드의 규칙 비교시 나올 수 있는 경우는 완전히 같은 경우, 완전히 다른 경우, 한쪽이 다른 쪽에 포함되는 경우, 두 규칙이 겹치는 경우의 네 가지가 있으나 앞의 두 번째 단계에서 대부분의 경우들이 고려되기 때문에 통합 알고리즘의 세 번째 단계에서는 두 노드의 규칙

### 5. 통합 알고리즘 구현

본 논문에서 제시한 분류체계의 통합 기법은 현재 SNUDCL project의 차기 project인 CatalogStop project에서의 한 모듈로 구현되었다. 자바를 이용하여 구현하고 있으며, 실제 각 쇼핑몰의 분류체계를 자료로 하여 테스트 작업을 진행 중이다.

### 6. 결론 및 향후 과제

본 논문에서는 규칙이라는 추가 정보를 가지는 인터넷 쇼핑몰의 분류체계들에 대하여 자동적인 통합을 위한 기법을 제시하였다. 이러한 기법으로 인해 쇼핑몰 사용자들에게 모든 쇼핑몰의 통합되고 정형화된 뷰의 제공, 사용자별 분류체계의 생성, 메타 쇼핑몰 간의 통일된 인터페이스 제공 등을 할 수 있음으로 인해서 전자상거래에서의 새로운 서비스 방향을 열었다는데 그 의의가 있다.

앞으로는 좀 더 자연스러운 분류체계의 통합과 보다 폭넓은 활용을 위하여 단순한 and, or로 표현되는 규칙이 아니라 fuzzy이론이 결합된 규칙에서도 잘 적용될 수 있게 보장이 이루어 질 것이다.

### 7. 참고 문헌

- [1] Jihye Jung, et al., EE-Cat: Extended Electronic Catalog for Dynamic and Flexible Electronic Commerce, IRMA, 2000
- [2] 정지혜, 이상구, 우치수, 전자상거래에서의 체계적인 상품 카탈로그 구축을 위한 분류 체계 모델 및 구현, 99 한국 데이터베이스 학술대회, 1999
- [3] Daniela Florescu, et al., Database Techniques for the World-Wide Web: A Survey, SIGMOD Record 27(3), 1998
- [4] Atsuyuki Morishima, et al., InfoWeaver: Dynamic and Tailor-made Integration of Structured Documents, Web, and Databases, ACM DL, 1999