

홈쇼핑 사이트의 정보를 검색하기 위한 온톨로지 설계

구미숙^o 황정희 류근호

충북대학교 데이터베이스 연구실

{gumisug^o, jhhwang, khryu}@dblab.chungbuk.ac.kr

Ontology Modelling for the Information Retrieval of Home Shopping Sites

MiSug Gu^o JeongHee Hwang KeunHo Ryu

Dept. of Computer Science, Chungbuk National Univ.

요 약

현재의 웹은 사용자가 원하는 정보를 정확하고 빠르게 검색 결과를 보여주지 못하는 단점이 있다. 그러므로 사용자에게 정확한 정보 전달을 해 주고자 시맨틱 웹이 등장하게 되었다. 시맨틱 웹은 기계가 이해할 수 있는 온톨로지를 구성하여 사용자가 원하는 정보를 정확하게 전달해 줄 수 있다는 점에서 미래의 웹으로 각광을 받게 될 것이다. 시맨틱 웹의 기반이 되고 있는 온톨로지는 어떤 특정 도메인에서 사용되는 정보들과 그 정보들 간의 관계를 정의해 놓은 것으로, 관련 도메인 전문가들과 협의에 의하여 개념들과 관계들의 구조를 정하고, 이를 기반으로 구축된다. 실제의 응용 시스템에서는 도메인마다의 구체적인 지식을 포함하는 온톨로지 설계가 필요하다. 이 논문에서는 택배회사가 홈쇼핑사이트 업체를 대상으로 효율적인 마케팅을 하기 위해 홈쇼핑사이트에 대한 기본정보를 추출하는 것을 목적으로 한다. 온톨로지를 구축하는 온톨로지 언어에는 RDF, RDF(S), DAML+OIL, OWL, Topic Map 등이 있다. 이 논문에서는 토픽맵을 사용하여 홈쇼핑 사이트 정보를 검색하기 위한 홈쇼핑 사이트에 대한 온톨로지를 설계하였다..

1. 서 론

온톨로지란 어떤 특정 도메인에서 사용되는 정보들과 그 정보들 간의 관계를 정의해 놓은 것으로, 관련 도메인 전문가들과 협의에 의하여 개념들과 관계들의 구조를 정의하여 구축한다. 그리고 실제의 응용 시스템에서는 도메인마다의 특정 지식을 포함하는 온톨로지가 필요하다. 온톨로지를 구축하는 온톨로지 언어에는 RDF, RDF(S), DAML+OIL, OWL, Topic Map 등이 있다[1]. 이 논문에서는 토픽맵을 사용하여 온톨로지를 작성하였다. RDF는 자원 중심이지만 토픽맵은 주제 중심으로 개념이나 사물에 대하여 정형화된 명세를 생성하기 위한 모델을 제시해주는 장점이 있다[1]. 토픽맵은 토픽, 연관관계, 어커런스 등 세 가지 기본 요소로 구성된다. 토픽은 주제를 의미하고, 어소시에이션은 주제와의 연관 관계를 나타내고, 어커런스는 주제에 대한 리소스가 위치한 정보를 가리킨다[2]. 이 논문은 택배회사의 효율적인 마케팅을 위해 홈쇼핑 사이트의 기본 정보를 탐색하기 위한 온톨로지 구축을 목적으로 한다. 따라서 이 논문의 토픽맵을 이용한 온톨로지는 두 가지 토픽 타입으로 지역과 품목을 정의하여, "shopping_topicmap" 온톨로지를 구축한다. 지역은 홈쇼핑 사이트를 운영하는 업체의 실제 주소에 대한 정보를 의미하고, 품목은 쇼핑 사이트에서 취급하고 있는 상품 품목을 의미한다. 우리는 이러한 토픽들을 가지고 토픽맵을 형성하여 홈 쇼핑 사이트가 위치한 지역과 그 사이트에서 판매하는 품목을 기준으로

하여, 홈쇼핑 사이트의 주소, 전화번호, 취급 상품, URL, 사이트 이름 등의 관련된 메타 정보인 회사 정보를 추출해 내는 것을 목적으로 한다. 온톨로지는 기존의 택소노미(taxonomy)와 같은 구조를 가지고 있어, 사용자가 원하는 정보 검색을 위해서 계층구조를 따라서 검색하는 방식이다. 본 논문에서는 택배 마케팅을 위해서, 홈쇼핑 사이트를 조사하여 온톨로지를 구축하였다. 이렇게 구축된 온톨로지를 통해서, 어느 특정지역에 위치한 특정 품목을 취급하는 홈쇼핑 사이트의 회사정보를 검색한다. 예를 들어, "대전시 서구 둔산동에 위치하여 컴퓨터를 취급하고 있는 홈쇼핑 사이트를 검색하십시오" 라는 사용자의 검색이 요청되었을 때, 기존의 키워드 검색의 결과는 사용자가 원하지 않는, 각 키워드에 해당하는 지역 정보나 컴퓨터 관련 정보를 다량으로 검색하는 단점이 있었다. 그러나 온톨로지를 이용한 정보 검색은 사용자가 필요로 하는 정보를 정확하게 전달해 준다.

2. 관련연구

[2]에서는, XTM(XML Topic Map)은 주제 중심으로 개념을 명세화하고 개념들 간의 연관 관계를 정의한 모델로서 ISO에서 표준으로 제안하고 있다. 초기에는 전자 색인을 위한 데이터 모델로 고안되었으나 현재에는 지식 관리 시스템의 지식맵, 컨텐츠 관리 시스템의 컨텐츠 맵 그리고 시맨틱 웹 온톨로지 등의 데이터 모델로 사용되고 있다. [3]에서, 토픽맵 모델의 세 가지 핵심 요소인 토픽, 연관관계, 어커런스에 대한 설명과 관계에 대해서 "opera" 토픽맵을 예를 들어 설명하고 있다. [6]에서는, 토픽맵을 이용하여 교수, 학생, 논문간의 관계를 정의 하여 구축된 온톨로지를 이용하여 검색을 좀 더 용이하게 하는 지식맵을 구현하였다. [4]에서는, 불경 버전

이 연구는 한국전자통신연구원의 정보통신 서비스 연구단의 연구비 지원으로 수행되었음

에 대한 작성자, 버전, 시대, 장소간의 관계를 정의하여 토픽맵 온톨로지를 구축하여 검색시스템을 구현하였다. [5]에서는, 토픽맵 모델 기반의 온톨로지를 생성, 저장, 검색하는 온톨로지 관리 시스템인 K-Box를 구현하였다. 기존의 온톨로지 응용은 웹상의 정보검색에 적용하지 않았지만 이 논문에서는 온톨로지를 웹상의 정보 검색에 적용하였다.

3. 홈쇼핑 사이트 토픽맵 온톨로지 모델

이 논문에서 제안하는 온톨로지를 구성하는 토픽맵 모델(Topic Map Model)의 기본 요소는 토픽, 연관 관계, 어커런스 등으로 구성된다[2].

3.1 토픽(Topic)

토픽은 사람, 사물, 개념, 의미 등 실제 존재하는 것과 속성이나 의미 등이 될 수 있다. 특정 문서의 토픽은 그 문서의 작성자가 나타내고자 하는 주제를 표현할 수 있는 단어들로 구성된다. 토픽맵의 토픽들을 비슷한 유형끼리 분류해서 토픽 타입으로 만든다. 토픽 타입들 간에도 상하관계와 상속관계가 존재한다[5]. 이 논문에서는 품목, 지역이라는 두 가지 토픽타입을 정의하였다.

3.2 어커런스(Occurrence)

각 토픽이 참조하는 자원과 연결된 정보를 어커런스라고 한다. 예를 들어, 홈 쇼핑 사이트에서 취급하는 컴퓨터에 대한 상품 정보를 나타내는 자원의 URL은 http://dblab.chungbuk.ac.kr/~computer을 참조한다. 어커런스는 문서파일, 이미지 파일, 데이터베이스 내 특정 레코드 등의 형태가 된다. 토픽맵의 토픽들은 자원을 가리키기 위해 HyTime, Xpointer 기법을 사용한다[6].

3.3 연관관계(Association)

토픽 연관관계는 둘 이상의 토픽들 사이의 상하관계 뿐 아니라 의미적인 관계를 정의한다.

"Daum selling Computer is located in Seoul."
 "Taejon is a subclass of Regions."
 "Items is a superclass of Cosmetic."

위의 예에서, 두 토픽인 "Daum"과 "Seoul"사이에는 "located_in" 관계, "Taejon"과 "Region" 토픽은 상하관계가 있고, 연관관계도 연관관계 타입이 있어 "located_in", "superclass", "subclass" 가 있다.

```
<?xml version="1.0" encoding="UTF-8" ?>
<topic id="tv">
  <instanceOf>
    <topicRef xlink:href="#home_electronic"/>
  </instanceOf>
  <baseName>
    <baseNameString>TV</baseNameString>
  </baseName>
  <occurrence>
    <resourceRef xlink:href="http://dblab.chungbuk.ac.kr/~tv"/>
  </occurrence>
</topic>
<association id="home_electronics-tv">
  <instanceOf>
    <topicRef xlink:href="#s2.1"/>
  </instanceOf>
  <member>
    <roleSpec>
      <topicRef xlink:href="#s2.2"/>
    </roleSpec>
    <topicRef xlink:href="#home_electronic"/>
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#s2.3"/>
    </roleSpec>
    <topicRef xlink:href="#tv"/>
  </member>
</association>
```

그림 1. XTM 이용한 온톨로지 문서

위의 그림1은 구축된 온톨로지 문서의 일부로서, 품목 인스턴스인 tv와 그 연관 관계를 표현한다. 토픽 타입과 연관관계 타입은 지식 및 정보 표현, 분류, 구조화를 위한 토픽맵의 중요한 기능이다. 토픽맵 표준안 스펙은 지금까지 살펴본 토픽, 어커런스, 연관관계 이외에도 여러 가지 지식을 관리하고 구조화하기 위한 요소가 있다.

4. 잠재 고객 검색을 위한 시맨틱 웹 온톨로지 설계

4.1 온톨로지 설계 방법

홈쇼핑 사이트 온톨로지를 설계하는 과정은 다음과 같은 단계를 거쳐서 온톨로지가 생성된다.

첫째, 온톨로지를 설계 하는 목적 및 범위를 설정한다. 이 논문에서는 택배 회사 마케팅을 위한 홈쇼핑 사이트의 정보를 효율적으로 추출하는 것이다.

둘째, 관련 개념을 추출하여 용어를 정의한다. 각각의 개념은 클래스가 되고, 각 개념마다 가지고 있는 속성을 파악하여 정의한다. 다음 그림2는 이 논문에서 정의한 품목과 지역 토픽과 회사정보의 속성을 정의한 것이다.

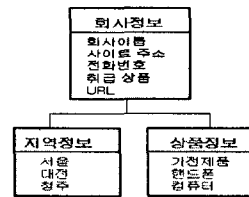


그림 2. 개념 속성정의

셋째, 개념 계층도와 개념 연관도를 작성한다. 계층 관계는 클래스와 그 속성에 의한 상속관계와 집합관계를 표현한다. 그리고 연관관계는 지식의 연관성을 기준으로 특정 지식을 중심으로 다른 지식과의 관계가 정의되고, 지식들 사이의 연결 경로는 그래프 구조로 표현한다.

아래 그림3은 이 논문의 연관관계 계층도를 표현하였다.

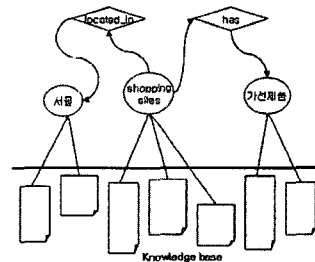


그림 3. 연관관계 계층도

넷째, 앞 단계를 기반으로 XTM 문서를 작성한다. 다섯째, 작성된 XTM문서에 대한 유효성에 대한 평가를 하여 오류나 누락된 부분이 있으면 수정을 한다. 이 논문에서는 유효성 검사를 Ontopia사의 토픽맵 툴인 Omnigator를 이용하였다.

여섯째, 기존의 토픽맵이 존재하는 경우 토픽맵 통합과정을 거쳐 토픽맵 온톨로지를 작성한다.

일곱째, 작성된 온톨로지는 XML 문서이므로 SaxParser

와 TmpParser에 의한 파싱과정을 거치게 된다. 파싱을 마친 문서는 TmBuilder에 의해 토픽맵이 생성된다. 여덟째, 이렇게 생성된 토픽맵은 XTM 1.0 DTD의 엘리먼트인 Topic, Scope, TopicMap, BaseName, Variant, VariantName, Association, Member, Occurrence등으로 토픽맵 구현 클래스를 생성하여 온톨로지 데이터 베이스에 저장된다. 다음 그림3은 토픽맵 생성 과정이다.

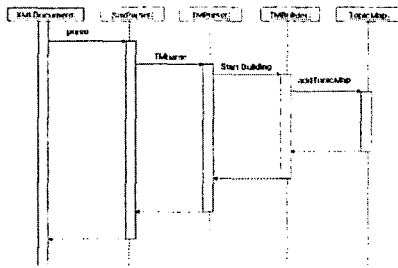


그림 4. 토픽 맵 생성 과정

4.2 홈쇼핑 사이트 정보 검색 위한 온톨로지 설계

시맨틱 웹의 온톨로지 기반의 홈쇼핑 사이트에 대한 검색 시스템은, 3장에서 소개한 토픽 맵 모델의 기본 요소인 토픽, 연관관계, 어커런스 등을 이용하여 온톨로지를 설계하였다.

첫째, "shopping_topicmap"을 작성하기 위해서 두가지 토픽 타입을 정의하여 토픽에 대한 설계를 하였다. 아이템 토픽 타입은 각 홈쇼핑 사이트를 조사하여, 각 사이트에서 취급하고 있는 품목들을 정의한다. 판매하고 있는 품목에는, 핸드폰, 가전제품, 컴퓨터, 의류등이 있다. 각 품목에 해당되는 구체적인 상품명이나 상품에 대한 정의는 이 논문의 범위를 벗어나므로 정의하지 않았다. 지역토픽은 쇼핑 사이트가 소재한 지역 정보를 의미하고, 각 도와 특별시, 광역시 등으로 구분하여 정의하였다. 그리고 그 하부 토픽으로 각 시와 도에 속하는 구, 군 등으로 토픽을 설계한다. 다음 그림4는 토픽간의 계층 구조를 표현 하였다.

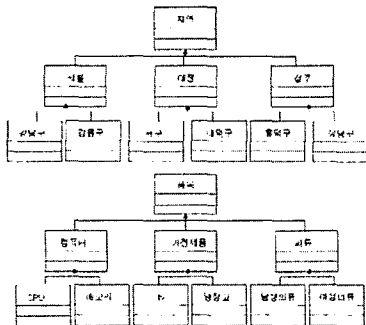


그림 5. 두 토픽의 계층구조

둘째, 두 토픽간의 연관관계를 정의 하였다. 지역, 품목이라는 토픽 타입은 각각 지역과 품목을 최 상위 클래스로 하고 서브 클래스를 가지고 있다. 지역 클래스는 각 도와 특별시, 직할시 등의 서브 클래스를 가지며, 각 구나 군이라는 서브 클래스를 가진다. 그리고 품목 클래스

는 가전제품, 의류, 컴퓨터 관련 제품 등의 세부적인 품목을 서브 클래스로 가진다. 그리고 회사와 지역 토픽 간에는 회사가 특정 지역에 위치한다는 "located_in" 연관관계가 있으며, 회사와 품목 토픽 간에는 회사가 품목을 가지고 있다는 관계인 "has" 관계가 있다. 다음 그림5는 토픽과 연관관계를 표현한 온톨로지 모델이다.

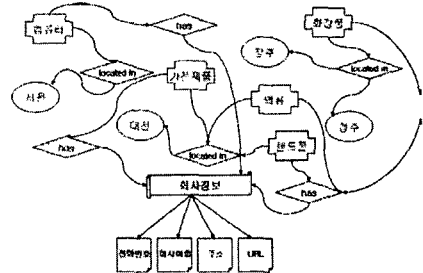


그림 6. 토픽맵 온톨로지 모델

두 가지 토픽 타입으로 "shopping_topicmap" 온톨로지를 구축하여 온톨로지 정보를 가지고 웹에 대한 검색을 한다. 앞서 질의를 했던 대전 둔산동 지역의 컴퓨터를 취급하는 업체에 대한 정보를 택배 회사에 제공해 주므로써 마케팅에 효율을 기할 수 있다.

이와 같이 XTM이라는 토픽맵을 이용하여 구축된 온톨로지는 메타데이터를 구조화 시켜서 기술하고, 토픽들간의 연관관계를 기술함으로써 사용자의 질의에 좀 더 정확한 정보 검색 결과를 가져다주는 시맨틱 웹의 기반이 된다.

5. 결론

이 논문에서는 택배회사 마케팅을 위해서 정보를 검색하기 위해서 시맨틱 웹을 이용하여, 지금까지의 웹의 키워드 검색보다 정확한 검색을 하기 위한 목적으로 온톨로지의 구축에 대한 설계를 하였다. 지금까지의 웹이 단순히 검색 결과 페이지를 디스플레이 하는 정도라면, 시맨틱 웹은 이와 같은 단점을 해결하고 사용자에게 정확하고 효과적인 검색 결과를 가져다주는 방법이다. 현재 우리는 이와 같은 시맨틱 웹에 의한 검색 시스템을 구현하고 있고, 앞으로 온톨로지를 자동으로 생성하기 위해 데이터 마이닝 기법을 적용할 예정이다.

참고문헌

[1] 김정민, 박철만, 정준원, 이한준, 민경섭, 김형주, "K-Box: 토픽맵 기반의 온톨로지 관리 시스템", 한국정보과학회, 2004
 [2] S. Pepper, B. Moore, "XML Topic Maps(XTM) 1.0", TopicMaps.Org.
 [3] Steve Pepper, "The TAO of Topic Maps", XML Conference & Exposition, 2000
 [4] Kounq-lung Lin, Yen-jen Oyang, "Knowledge Management for a Buddhism Digital Archive with Topic Map", ICDAT, 2002
 [5] 김정민, 박철만, 정준원, 이한준, 정호영, 민경섭, 김형주, "온톨로지 기반의 지식맵 서비스 시스템의 설계 및 구현", 정보과학회 춘계학술대회, 2003
 [6] 정호영, 김정민, 정준원, 김형주, "XTM 기반의 지식맵", 데이터베이스연구회지, 2003