데이터 마이닝 기법을 이용한 XML 문서의 온톨로지 반자동 생성

구 미 숙[†] · 황 정 희[††] · 류 근 호[†††] · 홍 장 의[††††]

요 약

최근 웹 문서를 비롯한 공공 문서 등에 대한 문서 교환을 위해 XML 데이터를 이용한 표준화 작업이 진행 중이므로 XML 문서가 증가하고 있다. 이와 같은 XML 문서에 대한 정보 검색의 효율을 높이기 위해 의미적 요소를 추가한 온톨로지를 기반으로 하는 시맨틱 웹이 등장하였다. 그러나 기존의 수동적인 온톨로지 구축 방식은 비용과 시간이 많이 소모되는 단점이 있으므로 이 논문에서는 유사한 도메인의 XML문서 집합으로부터 데이터 마이닝 기법의 연관규칙 알고리즘을 이용하여 반자동으로 온톨로지를 구축하는 방법을 제안한다. 제안한 방법은 특정한 도메인에 대한 온톨로지를 구축하기 위해서 필요한 데이터의 형태 및 개념 레벨, 그리고 얼마나 많은 개념을 사용할 것인가 하는 도메인 범위의 자동 설정을 온톨로지 자동 생성을 위한 온톨로지 도메인 레벨을 결정하기 위해서 데이터 마이닝 알고리즘을 이용한다. XML 문서의 태그에 대해 연관규칙을 적용하여 빈발하게 발생하는 빈발 패턴을 찾아내고, 서로 관련 있는 개념의 쌍을 추출하여 온톨로지 자동 생성을 위한 도메인 범위를 설정한다. 온톨로지 구축은 온톨로지 언어중의 하나인 XML Topic Maps와 공개 소스인 토픽맵 엔진인 TM4J를 이용하여 온톨로지 기반의 시맨틱 웹 엔진을 구현하였다.

키워드 : 온톨로지, 시맨틱 웹, 데이터 마이닝, XTM, TM4J

Semi-Automatic Ontology Generation about XML Documents using Data Mining Method

Mi Sug Gu[†] · Jeong Hee Hwang[††] · Keun Ho Ryu[†††] · Jang-Eui Hong[††††]

ABSTRACT

As recently XML is becoming the standard of exchanging web documents and public documentations, XML data are increasing in many areas. To retrieve the information about XML documents efficiently, the semantic web based on the ontology is appearing.

The existing ontology has been constructed manually and it was time and cost consuming. Therefore in this paper, we propose the semi-automatic ontology generation technique using the data mining technique, the association rules. The proposed method solves what type and how many conceptual relationships and determines the ontology domain level for the automatic ontology generation, using the data mining algorithm. Appying the association rules to the XML documents, we intend to find out the conceptual relationships to construct the ontology, finding the frequent patterns of XML tags in the XML documents. Using the conceptual ontology domain level extracted from the data mining, we implemented the semantic web based on the ontology by XML Topic Maps (XTM) and the topic map engine, TM4J.

Key Words : Ontology, Semantic Web, Data Mining, XML Topic Maps, TM4J

1. 서 론

현재 인터넷 기술의 발전으로 인해서 사용자들은 많은 다양한 정보를 접하고 있으나, 현재의 정보 검색 기법으로는 사용자가 원하는 정보를 정확하게 제시하지 못하기 때문에, 이와 같은 단점을 해결하기 위해 시맨틱 웹(semantic web)이 대두되고 있다. 시맨틱 웹은 온톨로지를 기반으로 하며, 온톨로지(ontology)는 특정 분야의 지식을 표현하기 위한 기본지식 체계를 제공하고 정보 검색에 대한 정확한 결과를 제시 하도록 도움을 준다. 그리고 XML이 현재의 웹 문서와 공공문서 등에서 표준화의 수단으로 이용되고 있기 때문에 많은 분야에서 XML문서가 증가하고 있다.

기존의 온톨로지는 전문가에 의해서 생성되었는데, 온톨로지 생성 시에 많은 데이터로 인해서 온톨로지 생성에 시간과 비용이 많이 소요된다. 그러므로 이 논문에서는 데이터 마이닝 기법을 이용한 반자동 온톨로지 생성 기법을 제안하며,

※ 이 연구는 산업자원부 한국산업기술평가원 지정 청주대 정보통신 연구센터 및 정보통신부 대학 IT연구센터 육성지원사업에 연구비지원으로 수행 되었음.
† 준 회 원 : 충북대학교 전자계산학과 박사과정
†† 정 회 원 : 남서울대학교 컴퓨터학과 전임강사
††† 종신회원 : 충북대학교 전기전자 컴퓨터공학부 교수
†††† 정 회 원 : 충북대학교 전기전자 컴퓨터공학부 교수
논문접수 : 2005년 5월 25일, 심사완료 : 2005년 12월 12일

이 기법은 사용자의 특정 도메인 내의 정보검색에 대하여 사용자에게 효율적인 정보를 제공하는데 도움을 준다.

기존의 온톨로지는 주로 텍스트 문서를 이용하여 구축하였으나, XML 문서가 증가 하므로 XML문서를 기반으로 하는 온톨로지 구축 방법의 연구가 필요하다. 그러므로 이 논문에서는 현재 웹 데이터의 표준인 XML 문서로부터 온톨로지를 생성하는 방법을 제안한다.

이 논문의 구성은 다음과 같다. 제 2장에서는 온톨로지와 온톨로지를 표현하는 XML Topic Maps와, 마이닝 기법을 이용한 기존의 온톨로지 생성 기법에 대해서 알아본다. 제 3장에서는 이 논문의 데이터로 사용하고 있는 XML 문서에 대하여 연관 규칙 알고리즘을 적용하는 방법 및 온톨로지의 도메인 레벨을 생성하는 과정에 대해서 알아본다. 제 4장에서는 XTM과 공개 소스인 토픽맵 엔진 TM4J를 이용한 온톨로지 자동 구축에 대해서 알아본다. 제 5장에서는 제안 시스템의 구현 환경과 연관규칙 알고리즘을 적용하여 온톨로지의 도메인 레벨의 생성 과정에 대한 실험 평가를 제시한다. 제 6장에서는 이 논문의 결론 및 향후 연구 방향에 대해서 기술한다.

2. 관련 연구

온톨로지는 특정 도메인의 개념 및 지식을 명세화하기 위해 그 지식을 설명하는 표준 용어들을 정의하고, 용어들 사이의 계층(taxonomy)및 연관 관계를 정의하는 것이다. 온톨로지는 개념과 관계로 구성되는데, 개념은 온톨로지 내에 정의된 키워드를 의미하고, 관계는 온톨로지 내에 정의된 개념간의 관계를 의미 한다[1-3, 21, 22].

그리고 XML Topic Maps(XTM)는 주제 중심으로 개념을 명세화 하고 개념들 간의 연관 관계를 정의한 모델로서 ISO의 표준안으로, 초기에는 전자 색인을 위한 데이터 모델로 고안되었으나 현재는 지식 관리 시스템의 지식 맵, 컨텐츠 관리 시스템의 컨텐츠 맵 그리고 시맨틱 웹 온톨로지 등의 데이터 모델로 사용되고 있다[23].

토픽맵 모델의 세 가지 핵심 요소는 토픽(Topic), 어커런스(Occurrence), 연관관계(Association)등이다. 토픽은 주제, 어소시에이션은 주제와의 연관 관계, 어커런스는 주제에 대한 리소스가 위치한 정보를 가리키는 역할을 한다[4,5]. 그리고 [24]에서는 토픽맵을 이용하여 교수, 학생, 논문간의 관계를 정의하여 온톨로지를 구축하여, 검색에 효율을 기한 지식맵을 구현한 예를 보여준다. 그리고 [6]은 불경의 버전에 대한 작성자, 시대, 장소간의 관계를 정의하여 토픽 맵 온톨로지를 구축하여 검색시스템을 구현하였다. [23]은 토픽맵 모델을 기반으로 온톨로지를 생성, 저장, 검색하는 온톨로지 관리 시스템인 K-box를 구현하였는데, 온톨로지 관리를 위한 기본적인 기능을 제공하며, 이질적인 저장소들을 일관된 인터페이스로 접근할 수 있게 한다.

데이터 마이닝 알고리즘은 많은 양의 데이터로부터 유용한 정보를 추출해 내는 방법이다. 최근 반 구조화된(semi-Structured) 문서로서 표준화의 기본이 되고 있는 XML문서에 대한 데이터 마이닝 기법의 적용 연구가 시도되고 있다[7-12].

[13]에서는 텍스트 데이터에 대한 언어 처리 과정을 통해서 생성된 텍스트문서에서, non-taxonomic 개념관계를 찾아내는 새로운 접근법을 제시하였다. 개념간의 관계를 찾을 뿐 아니라, 관계를 정의할 수 있는 적절한 추상화된 레벨를 결정하기 위해 일반화된 연관 규칙을 사용한다. 특정한 온톨로지 내에서, 개념 관계를 얼마나 많이 그리고 어떠한 타입으로 모델링 할 것인가에 대하여, 텍스트 데이터를 기반으로 일반화된 연관 규칙을 적용하여 개념관계를 필터링 하는 과정을 수행하였다.

[14]에서는 비용을 적게 들이고 짧은 시간이내에, 대량의 적절한 온톨로지를 생성하기 위한 개념구조를 찾아내어 일반적인 아키텍처를 제시하고 있다. 이 아키텍처를 기반으로, 반자동화된 택소노미(Taxonomy) 구조를 만들어 내는데, 개념간의 관계를 찾을 뿐 아니라, 관계를 정의할 수 있는 적절한 추상화된 레벨을 결정하기 위해서 일반화된 연관 규칙을 사용한다. 이 알고리즘은 조상 노드를 포함시켜서 계층 구조 생성을 위한 추상화 레벨을 일반화 시켜가는 과정으로 전개된다.

기존의 수동적인 온톨로지 구축방법은 비용과 시간이 많이 소모되는 단점이 있으므로 이 논문에서는 유사한 도메인의 XML문서 집합에 대한 데이터 마이닝 알고리즘을 이용하여 반자동의 온톨로지를 구축하는 방법을 제안한다.

3. 연관규칙을 이용한 반자동 온톨로지 도메인 레벨 생성

3.1 토픽 맵 온톨로지 모델

이 논문에서 제안하는 온톨로지를 구성하는 토픽맵 모델의 기본 요소는 토픽, 연관 관계, 어커런스 등으로 구성된다[4,5].

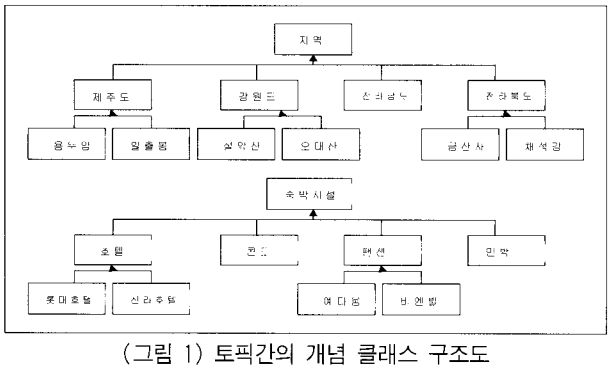
토픽은 사람, 사물, 개념등 실제 존재하는 것과 속성이나 의미 등이 된다. 특정 문서의 토픽은 그 문서의 작성자가 나타내고자 하는 주제를 표현할 수 있는 단어들로 구성된다. 토픽맵의 토픽들은 비슷한 유형끼리 분류한 토픽 타입과 상하관계와 상속관계가 존재한다[23]. 이 논문에서는 온톨로지 생성 도메인을 관광정보로 제한하며, 온톨로지를 생성하기 위하여 지역, 관광지, 숙박시설, 교통시설, 토속음식, 관광 상품 등을 토픽타입과 토픽으로 정의한다.

어커런스는 각 토픽이 참조하는 자원과 연결된 정보를 의미하고, 각 토픽은 자신이 참조하는 하나 이상의 실제적인 지식 항목과 연결되어 있다. 예를 들어, http://dblab.chungbuk.ac.kr/~tour은 관광에 대한 정보를 연결시켜 자원을 참조한다. 어커런스는 문서파일, 이미지 파일, 데이터베이스 내 특정 레코드 등의 형태가 된다. 토픽맵의 토픽들은 자신의 자원을 가리키기 위해 하이타임(HyTime), 엑스 포인터(Xpointer)

기법을 사용한다. 토픽맵과 실제 지식 자원 사이를 분리할 수 있는 것은 실제 지식 항목의 이동 없이 토픽맵 만으로 지식 항목을 분류하는 색인 기능을 제공한다[24].

토픽의 연관관계는 둘 이상의 토픽들 사이의 상하관계뿐 아니라 의미적인 관계를 정의한다. 예를 들어 보면, "Yongduam is located in Chejudo.", "Jeju Pacific Hotel is one of the accommodations." 위의 예에서, 두 토픽인 "Yongduam"과 "Chejudo"사이는 "be located in" 관계, "Jeju Pacific Hotel"과 "accommodations" 사이에는 "is one of" 관계가 있다. 토픽은 토픽 타입, 연관관계는 연관관계 타입, 상하관계인 "superclass", "subclass"등의 관계가 있다. 토픽 타입과 연관관계 타입은 지식 및 정보 표현, 분류, 구조화를 위한 토픽맵의 중요한 기능이다.

아래(그림 1)은 온톨로지 구축을 위해서 각 토픽간의 개념 관계를 계층구조로 표현한 클래스 구조이다. 지역에 대한 클래스 계층구조는 지역을 루트로 제주도, 강원도, 전라남도, 전라북도 등을 서브 클래스로, 그리고 각 도의 관광지 이름 등이 서브 클래스이다. 그리고 숙박시설이 루트인 클래스 구조는 각 숙박 시설인 호텔, 콘도, 팬션, 민박 등이 서브 클래스이며, 그 아래에는 각 호텔, 콘도, 팬션 등의 이름이 서브 클래스가 있다.



(그림 1) 토픽간의 개념 클래스 구조도

3.2 관광 정보를 표현하는 XML 문서의 예

기존의 텍스트 데이터를 이용한 온톨로지 생성은 텍스트가 계층구조를 가지고 있지 않기 때문에, 택소노미 구조를 만들기 위해서 일반화된 연관규칙 알고리즘을 이용하여, 빈발한 개념의 쌍의 조상 노드를 찾아가는 방식으로 마이닝을 수행한다[12, 13, 14, 15].

그러나 XML문서는 태그가 계층구조(hierarchical structure)와 일반화된 특성이 있으므로 XML 문서에 대한 연관규칙 알고리즘을 적용하여 데이터 마이닝 과정을 수행하여, 빈발하게 발생하는 빈발 패턴을 찾아낸다. 이 논문의 데이터는 웹에 있는 관광에 대한 정보를 제공하는 사이트를 통해서 관광 관련 자료를 Xgenerator[16] 라는 문서 변환기를 이용하여 XML 문서로 전환하는 과정을 수행한 XML 문서를 이용하였다.

XML 문서에 데이터 마이닝 알고리즘을 적용하여, XML 문서의 특징과 태그를 기반으로 조인 과정을 통해서 태그의 빈발 패턴을 찾아내어 서로의 연관성에 대해 사용자의 질의에 적합한 온톨로지를 구축하여 검색 결과를 제시한다. 예를 들어, "제주도에 있는 용두암을 여행할 때 적절한 숙박시설을 검색하시오"라는 사용자의 질의를 가정하고, 사용자 질의에 대해서 XML 문서의 태그인 다음의 관계를 보자.

<관광지이름>용두암</관광지이름> ⇒ <호텔>제주 퍼시픽 호텔</호텔>의 관계를 찾기 위한 데이터 마이닝을 수행한다. XML 문서의 태그에 대한 빈발 패턴을 찾아내어 연관관계를 맺어주는 연관 규칙을 수행하여 온톨로지의 도메인을 생성한다. 다음 (그림 2)는 온톨로지 구축을 위한 도메인으로 설정한 관광 정보에 관한 내용을 XML로 전환한 문서의 일부분을 보여주고 있으며, tour_info를 루트 태그로 해서 지역, 관광지, 숙박시설 등의 서브 태그 구조를 가지고 있다.

```
<?xml version="1.0" encoding="euc-kr"?>
<Tour_Info>
        <지역><국내><도시=제주도>
        <관광지>
                <위치>제주시</위치>
                <관광지이름>용두암</관광지이름>
        </관광지>
        <숙박시설>
                <호텔>
                        <호텔이름>Jeju Pacific Hotel</호텔이름>
                        <객실>
                                <객실종류>스탠다드</객실종류>
                                <객실종류>디럭스더블</객실종류>
                        </객실>
                        <주소>제주시 용담동</주소>
                        <전화번호>064-758-2500</전화번호>
                        <쇼핑>토산품점</쇼핑>
                        <음식점>한식당</음식점>
                        <음식점>양식당</음식점>
                </호텔>
                <팬션>
                        <팬션이름>예다움팬션</팬션이름>
                        <객실>
                                <객실종류>12평형 원룸</객실종류>
                                <객실종류>16평형 원룸</객실종류>
                        </객실>
                        <주소>제주시 용담동</주소>
                        <전화번호>064-742-4938</전화번호>
                </팬션>
        </숙박시설>
        </도시></국내></지역>
```

(그림 2) 관광정보에 대한 XML 문서

3.3 온톨로지 자동구축을 위한 연관 규칙 알고리즘

이 절에서는 XML 문서에 대한 온톨로지 자동 구축을 위해서 사용하고 있는 연관 규칙 알고리즘에 대해서 설명한다. 먼저 연관규칙 알고리즘의 기본 원리에 대해서 살펴본다.

데이터베이스의 트랜잭션의 항목에 대한 $X_k \rightarrow Y_k$의 모든 가능한 연관규칙에 대한 신뢰도와 지지도를 계산하여 연관규칙에 의해서 나온 결과에 대하여 가지치기를 한다.

첫째. 항목들의 전체집합 I에서, 미리 결정된 최소 지지도인 *min_sup* 이상의 트랜잭션 지지도를 가지는 항목들의 모든 집합을 빈발 항목집합(large itemsets)이라고 한다.

둘째, 모든 빈발 항목집합 1의 모든 공집합이 아닌 부분집합을 찾는다. 각 부분집합 a에 대해서 *supp(a)*의 비율이 최소 신뢰도인 *min_sup* 이상이면 규칙을 출력한다[7, 14].

온톨로지 구축을 위해 사용할 데이터인 웹의 관광정보 사이트를 통해서 찾아낸 관광정보들에 대하여 데이터 마이닝을 하기 위한 전 처리 과정에 대해서 알아본다.

첫째, 웹에 있는 관광 정보 사이트를 대상으로 관광정보를 찾아내어 XML 제너레이터를 이용하여, XML 문서로 전환하는 과정을 수행한다. 둘째, XML 문서에 대해서 Dom Parser를 이용하여 파싱을 한다. 셋째, 관광 정보로 만들어진 XML 문서를 구성하고 있는 약 2500개의 태그를 추출한다. 넷째, 관광정보 사이트를 통해서 찾아낸 정보로 이루어진 XML 문서에서 추출해낸 태그에 대해서, 같은 의미를 나타내는 유사한 표현을 사용하는 태그 명을 찾기 위해 WordNet (http://www.cogsci.princeton.edu/~wn/wn2.0)을 이용하여 처리한다[17, 25]. 다섯째, XML 문서의 태그를 추출하여 연관규칙 알고리즘을 수행하기 위해서는 각 태그에 번호를 매겨서 매핑 테이블을 만들어 프로그램을 실행한다.

이와 같은 전 처리 과정을 수행한 XML 문서에 대해서 연관규칙을 이용하여, 온톨로지 도메인 레벨의 생성을 위한 개념관계를 찾아내기 위해서 다음과 같은 과정을 수행한다.

첫째, 트랜잭션을 결정한다. 관광정보의 XML 문서가 하나의 트랜잭션이 된다. 둘째, 항목을 결정한다. 모든 XML 문서에 있는 각 태그가 항목이 된다. 셋째, 모든 연관규칙에 사용자가 지정한 최소 지지도와 신뢰도를 결정한다. 넷째, 최소 지지도와 신뢰도를 초과하는 연관 관계를 찾아낸다. 다섯째, XML 문서의 태그간의 연관 관계를 찾아서 연관성이 없는 관계를 제거한다. 여섯째, 온톨로지 생성을 위한 개념 계층구조를 생성한다.

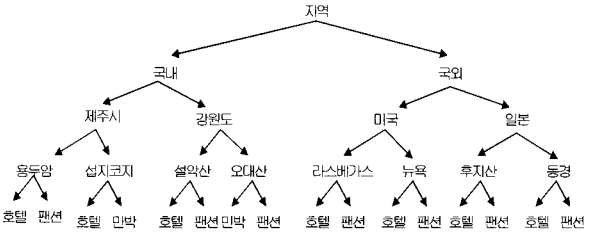
XML 문서내의 태그로 이루어진 일반화된 계층구조를 이용하여, 연관규칙 알고리즘을 적용한다. 관광지 정보, 숙박시설 정보, 교통 정보를 가지고 있는 문서가 연관규칙에 적용시킬 수 있는 하나의 트랜잭션이 된다. 이와 같은 각 문서 단위의 트랜잭션이 가지고 있는 태그를 항목으로 구성하며, 구성된 항목은 그 문서 내에서 발생 빈도를 계산한다.

3.4 온톨로지를 위한 도메인 레벨 생성 과정

온톨로지 도메인 레벨을 생성하기 위해서, XML 문서의 태그를 이용하여 일반화된 형태의 계층구조를 만들어 낸다.

```
<국내><지역= 제주도>
            <관광지><위치>제주시</위치>
            <관광지이름>용두암</관광지이름></관광지>
      </지역>
      <지역 = 강원도 >
      <관광지><위치>속초</위치>
      <관광지이름>설악산 국립 공원</관광지이름></관광지>
      </지역>
</국내>
<국외><지역= 미국>
            <관광지><위치>네바다</위치>
            <관광지이름>라스베가스</관광지이름></관광지></지역>
      <지역 = 일본 >
            <관광지><위치>시즈오카현</위치>
            <관광지이름>후지산</관광지이름></관광지>
       </지역>
</국외>
```

(그림 3) 관광지역 사이트에 대한 XML 문서



(그림 4) 관광지역 사이트에 대한 XML 문서의 층구조

〈표 1〉 XML 문서에 의해서 생성된 트랜잭션

TID	Items
A	국내, 국외, 지역, 관광지, 위치, 호텔, 민박, 팬션, 주소, 관광상품
B	숙박시설, 지역, 호텔, 팬션, 호텔이름, 국외, 객실, 주소, 국내, 객실종류
C	국내, 지역, 관광지, 위치, 토속음식, 음식점, 음식이름, 주소, 전화번호
D	국내, 지역, 관광지, 위치, 관광 상품, 관광 상품이름

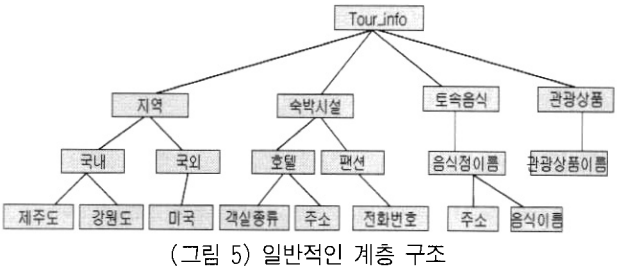
이 계층구조를 이용해서 온톨로지를 생성하고자 할 때, 대량의 데이터로 인해서 비용과 시간 면에서 비효율적이기 때문에, 온톨로지 구축을 위한 도메인 레벨을 위한 데이터 마이닝 기법인 연관규칙을 이용한다. 이와 같은 연관규칙을 사용하는 목적은 어떤 형태, 그리고 얼마나 많은 개념관계를 추출하느냐를 결정하기 위한 것이다. XML 문서의 태그간의 연관 관계를 찾는 학습을 수행하여, 연관성이 없는 관계는 제거하여 온톨로지 생성에 적당한 형태의 개념 계층구조를 만들어 낸다.

예를 들어, 다음과 (그림 3)의 태그로 이루어진 XML 문서를 이용하여 계층구조로 표현하면 그림 4로 나타낼 수 있다. 각 문서는 관광지에 대한 정보를 취급하는 사이트로부터 추출한 XML 문서의 일부분이며 테이블 1은 이것을 기반으로 생성된 트랜잭션이다.

위의 A, B, C, D 트랜잭션에서 지역, 주소, 호텔, 국내, 국외, 팬션 등이 네 가지의 트랜잭션에 함께 포함되어 있는 것을 확인할 수 있다. 이와 같은 빈발 패턴에 대한 연관 쌍을 추출하는 연관규칙 알고리즘을 반복해서 수행하여 빈발

〈표 2〉 개념에 대한 연관 쌍

Tag	Attribute	Tag	Attribute
국내	지역	Hotel	숙박시설
국내	관광지	음식점이름	토속음식
강원도	지역	관광지이름	관광지
관광 상품이름	관광 상품	관광지이름	관광지



(그림 5) 일반적인 계층 구조

패턴을 찾아낸다. 그리고 이 과정에서 함께 추출된 항목이 포함관계에 있는 경우에는 좀 더 일반화된 태그를 선택하여 온톨로지 도메인 레벨을 생성하는데 사용한다. 즉, 연관관계가 <호텔>과 <지역>이라는 쌍과, <숙박시설>과 <지역>이라는 쌍이 빈발 패턴으로 함께 추출되었을 경우 개념관계에서 일반화된 상위 태그인 <숙박시설>과 <지역>이라는 쌍을 선택한다.

XML 문서에서 서로 관련 있는 용어와 속성 값을 추출해 내기 위해, 관광 정보를 취급하는 사이트로부터 문서 집합을 추출한다. 각각은 관광 정보인 관광지, 숙박시설, 국내, 국외, 교통시설 등의 정보를 XML 문서로 전환한 문서집합으로 이루어져 있다. 이와 같은 문서에 대해서 언어적 처리 과정을 통해서 관련된 2,500개의 XML 태그를 추출한다. 알고리즘은 일반적인 개념간의 연관관계를 기반으로 일반화된 계층구조를 유도해내고, 온톨로지 생성을 위한 개념에 대한 적절한 관계를 찾아낸다.

다음 <표 2>는 XML 문서의 태그와 관련 있는 용어와 속성의 쌍을 학습한 내용을 표로 만든 것이며, (그림 5)는 XML 문서에서 개념의 연관관계를 찾아 온톨로지 구축을 위한 일반적인 계층를 표현한 것이다.

다음 단계는, 신뢰도와 지지도를 이용하여 그 비율이 낮은 개념관계 쌍은 온톨로지 생성에 적절하지 못하므로 제거한다.

아래 <표 3>은 이와 같은 연관 규칙 알고리즘을 실행하여 나온 결과로써, 개념관계 쌍의 연관관계를 찾아가는 과정에서 제거된 내용을 선을 그어서 표현한 것으로, (제주도, 관광지이름), (호텔이름, 호텔), (민박, 주소) 등의 개념관계는 제거되는 이유는 다음과 같다.

〈표 3〉 발견된 연관관계

Discovered relation	Confidence	Support
(국내, 지역)	0.37	0.04
~~(제주도, 관광지이름)~~	~~0.11~~	~~0.03~~
~~(강원도, 관광지)~~	~~0.39~~	~~0.04~~
~~(호텔이름, 호텔)~~	~~0.15~~	~~0.01~~
(민박, 주소)	0.22	0.02
(관광상품, 전라남도)	0.35	0.03
(관광지, 관광상품)	0.38	0.05

첫째, (제주도, 관광지이름)이라는 개념관계의 쌍에 대한 상위 관계의 개념인 (국내, 지역)의 쌍이 빈발 패턴으로 있기 때문에, 그 자식 관계에 있는 개념 쌍은 좀 더 일반화된 태그인 조상 태그에 포함 되므로 제거된다. 둘째, (호텔이름, 호텔)과 (민박, 주소) 쌍은 신뢰도와 지지도가 낮기 때문에 제거된다.

이와 같은 과정을 수행하여 연관규칙 알고리즘에 의해서 빈발하다고 판단되는 연관관계의 쌍을 찾아내어 온톨로지를 생성하는데 이용한다. 온톨로지를 구축할 때 포함관계에 의해서 상위 레벨의 지지도와 신뢰도가 높으면, 하위 레벨은 제거되는 과정을 거쳐서 일반화된 형태의 계층구조가 만들어진다. XML 문서가 가지고 있는 계층구조가 일반화 되어 있으므로, XML문서의 계층구조를 따라서 연관규칙을 적용하므로 써 일반화되어 있는 개념들을 선택하는 것이 가능하다. 그러므로 알고리즘에 의해서 적절하지 않은 태그간의 관계는 가지치기해서, 개념 관계를 설명하는 가장 적당한 온톨로지 도메인 레벨을 결정한다.

예를 들어 각 관광지마다 연관 관계를 맺을 수 있는 숙박시설인 호텔, 콘도, 팬션, 민박 등을 포함하는 XML 문서에서 {관광지} ⇒ {숙박시설} 관계나 {관광지} ⇒ {관광 상품} 등의 적절한 연관관계를 찾기 위해서, 연관규칙을 이용한 데이터 마이닝을 통해서 온톨로지 구축에 적당한 도메인 레벨을 결정 한다.

4. 온톨로지 자동구축

4.1 토픽맵 생성 과정

3장에서 연관규칙 알고리즘을 이용하여 생성된 빈발 태그 패턴과 XML 문서의 개념관계 쌍의 결과로 나온 태그 이외에도 빈발하게 발생하는 태그를 추가하여 생성한 온톨로지의 택소노미 구조를 XTM을 이용하여 만든다.

관광 정보에 대한 온톨로지의 구축은, XML문서의 태그인 "tour_info"를 루트로 시작한다. 추출된 개념관계 쌍인 관광지 정보와 숙박 시설에 대한 정보를 분류하기 위해서 먼저 국내와 국외로 구분하였다. 국내의 관광지중에서 제주도 지역에 있는 관광지 정보와 그 지역과 관련된 숙박시설에 대한 정보를 나타내고 있다. "국내", "제주도", "관광지", "용두암", "숙박시설", "제주 퍼시픽 호텔" 등이 토픽 맵의 토픽이 된다.

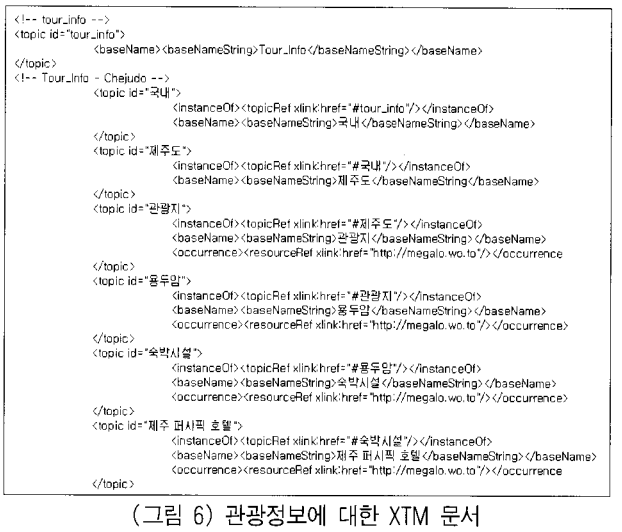
그리고 토픽맵의 어소시에이션은 "tour_info" 토픽이 루트가 되고, "국내" 토픽은 루트 "tour_info"의 서브노드가 된다. 그리고 "제주도" 토픽은 "국내" 토픽의 서브 노드이며, "관광지" 토픽은 "제주도" 토픽의 서브노드가 되고, "용두암"은 "관광지" 토픽의 서브노드이다. 그리고 "숙박시설"은 "용두암"의 서브노드가 되고, "제주 퍼시픽 호텔"은 "용두암"의 서브노드이다. 이와 같은 XTM 토픽맵의 구조를 이용하여 관광에 관한 정보를 표현한 XML 문서를 이용하여 온톨

로지를 구축하게 된다.

이 논문에서 온톨로지 구축을 위해서 이용한 XTM 온톨로지 구축 소스의 일부분을 아래 (그림 6)에서 볼 수 있다.

이와 같은 과정을 수행하여 구축된 온톨로지 문서는 XTM 문서이므로, 컨테이너인 자카르타-톰캣(Jakarta-Tomcat)을 실행 시켜 토픽맵 문서를 온토피아(Ontopia)사의 토픽맵 툴인 옴니게이터를 사용하여 유효성 검사를 하여[18], 하이버네이트를 이용하여 객체 관계형으로 온톨로지 데이터베이스에 저장한다[19].

그리고 유효성 검사를 한 타당한 문서인 경우, 온톨로지를 데이터베이스에 저장하기 위해서는 삭스 파서(Saxparser)와 토픽 맵 파서(TMParser)를 통해 파싱 과정을 수행하고, 토픽 맵 엔진인 TM4J에 있는 TMBuider를 이용하여 토픽맵을 자동으로 생성하게 된다. 기존의 토픽맵이 존재하고 있으면 추가나 업데이트 과정을 수행하여 토픽맵이 생성된다.



(그림 6) 관광정보에 대한 XTM 문서

4.2 TM4J를 이용한 온톨로지 자동 생성과정

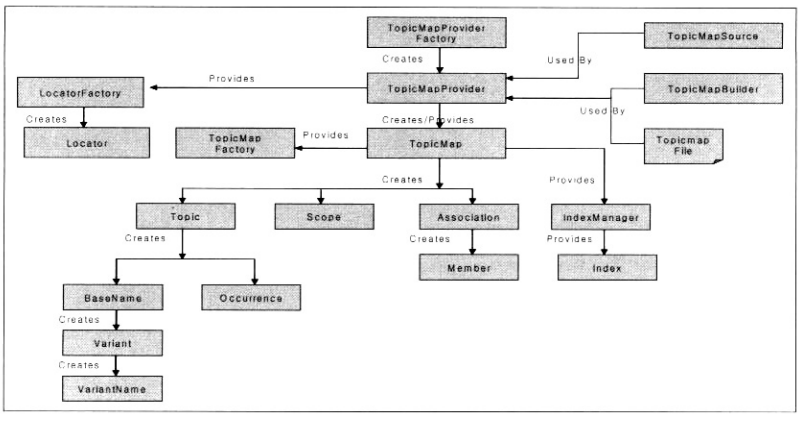
이 논문에서는, 토픽맵의 자동생성을 위해서, http://www.ontopia.net 사의 공개 소스인 토픽맵 툴 TM4J를 이용하였다. TM4J는 다음과 같이 구성되어 있다[18].

토픽 맵 객체들의 단일 인터페이스 제공을 위한 토픽맵 오브젝트 래퍼(Topic Map Object Wrappers), 토픽맵 객체들을 저장하기 위한 스토리지 래퍼(Storage Wrappers), 사용자가 요청하는 토픽맵 객체를 전달하는 토픽맵 제공자(Topic Map Provider), 대용량의 토픽맵의 효율적인 검색을 위한 토픽맵 캐쉬 관리자(Topic Map Cache Manager), 토픽맵 객체들을 생성하는 토픽맵 생성자(Topic Map Factory), 토픽맵 관리를 위해 일관된 인터페이스를 제공하는 토픽맵 관리자(Topic Map Manager), 토픽맵 가져오기(import), 내보내기(export) 등의 기능을 제공하는 토픽맵 도구(Topic Map Utils)등으로 구성된다.

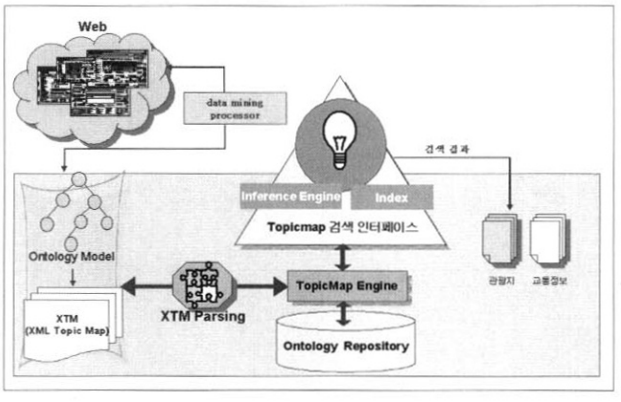
토픽맵 오브젝트 래퍼는 토픽맵을 구성하고 있는 오브젝트들의 구현에 비종속적으로 일관된 인터페이스를 제공해주는 역할을 하고, 자바의 인터페이스로 정의가 되며, 각각의 오브젝트들의 메소드를 정의하고 있다. 토픽맵 스토리지 래퍼는 토픽맵의 각 객체들을 실제의 스토리지에 영구적으로 저장하고 로드 하는 등의 작업을 수행하며, 토픽맵 오브젝트 래퍼를 통해서 하위의 스토리지에 관계없이 일관적인 인터페이스를 제공해 준다.

토픽맵 생성자는 토픽맵의 각 구성요소들을 생성하는 역할을 수행한다. 토픽맵 스토리지 래퍼별로 독립적으로 제공되며, 토픽맵 오브젝트 래퍼의 생성자 인터페이스를 통해 일관적인 인터페이스를 제공한다. 토픽맵 제공자는 스토리지에 저장된 토픽맵을 메모리로 로드 하여 사용자에게 제공하는 역할을 한다. 토픽맵 오브젝트 래퍼의 제공자 인터페이스를 통해 일관적인 인터페이스를 제공할 수 있다. 토픽맵 캐시 관리자는 특정 토픽을 검색하면서 그 토픽과 연관관계에 있는 다른 토픽들의 정보도 미리 검색하여 캐쉬에 저장하여 검색 조건에 부합되는 토픽이 있으면 디스크를 접근하지 않고 캐쉬에서 그 토픽을 제공하여 효율을 증가시킨다.

아래 (그림 7)은 TM4J의 기본적인 아키텍쳐이다. XTM으로 작성된 토픽맵 소스, 기존의 토픽맵 파일, 토픽맵 빌더(TopicMapBuilder)에 의해서 생성된 문서등이, 토픽맵 프로바이더 (TopicMapProvider)에 의해서 로케이터 팩토리



(그림 7) TM4J의 기본 아키텍쳐



(그림 8) 온톨로지 기반 검색 엔진 플랫 폼 구조도

(LocatorFactory)로 제공되면, 로케이터(Locator)가 만들어진다. 그리고 토픽맵 프로바이터 팩토리(TopicMapProvider Factory)가 토픽맵 프로바이더(TopicMapProvider)를 만들어서 토픽맵을 생성한다. 이렇게 생성된 토픽맵은 토픽 맵 팩토리에 제공되고, 인덱스 매니저에 제공되어 인덱스를 생성한다. 그리고 위에서 생성된 토픽맵은 토픽, 범위(scope), 어소시에이션을 만들어낸다. 토픽은 베이스 네임, 다른 이름(variant name), 어커런스를 생성하고, 어소시에션은 멤버를 생성한다. 그리고 그림 8은 온톨로지 기반의 검색 엔진 플랫 폼에 대한 구조도이다.

(그림 8)의 전체적인 흐름 과정에 대해서 설명하면, 웹 서버에 있는 HTML, XML, 기타 문서 파일들을 검색엔진에 의해서 가져온 자료들을 변형 과정을 수행하여 XTM 문서가 생성되면 파싱을 수행하게 된다. 그리고 온톨로지 기반 토픽맵 엔진의 처리 과정을 수행하여 온톨로지 데이터베이스에 저장되며, 사용자 인터페이스를 통해서 사용자의 질의에 알맞은 검색 결과를 제시해 준다.

5. 구현 및 검색 결과

온톨로지를 이용한 검색 엔진은 J2SDK 1.4.2(Java)와 공개 소스인 온토피아사의 토픽맵 엔진인 TM4J(Topic Map for JAVA)-0.9.6을 이용 하였고, 검색 인터페이스는 JSP와 HTML을 사용하여 구현하였다. 그리고 XML 데이터와 온톨로지 데이터의 저장, 관리를 위한 DBMS는 Oracle9i를 이용하였고, Java 컨테이너로 Tomcat 5.0을 사용하였다.

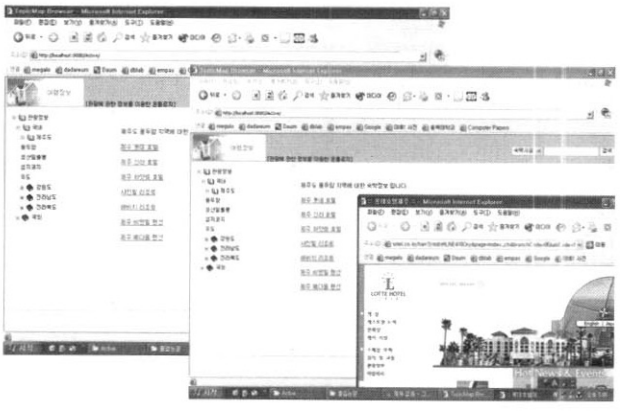
5.1 온톨로지를 이용한 정보 검색 시나리오

관광관련 사이트에 사용자들로부터 자주 발생하는, "제주도에 있는 용두암을 여행할 때 적절한 숙박 시설을 검색하시오." 와 같은 질의에서, "용두암"은 "관광지이름"중의 하나이고, "호텔"은 "숙박시설"중의 하나이다. 데이터베이스에 저장된 XML 문서에 대해서 사용자의 질의를 기반으로 서로 관련된 개념간의 연관관계를 찾아낸다.

사용자에 의해서 정보 검색 질의가 제기 되었을 때, XPath를 기반으로 XQuery의 신택스 구조[20, 26-28]를 이용하여

```
FOR $p IN DISTINCT
        doc(tour.xml)/지역/제주도/관광지이름
        RETURN
        <RESULT>
                $p,
                FOR $a IN DISTINCT /지역/제주도[관광지
                                        이름=$p]/숙박시설
                        RETURN $a
        </RESULT>
```

(그림 9) XQuery를 이용한 질의



(그림 10) 온톨로지 이용한 검색 결과

XML 문서에 대한 연관 규칙 알고리즘을 적용하여 데이터 마이닝을 수행한다. (그림 9)는 앞에서 제시한 질의를 XQuery를 이용한 것이다.

위와 같은 XQuery를 이용한 질의가 사용자에 의해서 제기 되었을 때, 제주도 지역의 관광지 이름과 숙박 시설에 대한 검색 결과를 제시하게 된다. 이와 같은 질의에 대한 온톨로지 계층 구조를 이용한 검색결과 화면은 (그림 10)에서 볼 수 있다.

왼쪽 메뉴에는 관광지를 의미하는 지역별로 구분되어 있어서, 계층구조의 메뉴를 클릭해서 지역에서 제주도를 선택하고, 제주도의 관광지 중에서 용두암을 선택한다. 그리고 오른쪽 상단에 있는 키워드 검색 창을 통해서 사용자가 원하는 숙박정보, 교통정보, 관광 상품 정보 등을 선택하여 클릭하면, 오른쪽 하단에 사용자가 원하는 정보가 제공된다. 그리고 화면에 띄워진 URL을 선택하면 오른쪽 하단에 보이는 작은 창으로 관련된 정보가 들어있는 웹 사이트로 연결된다.

5.2 실험 및 결과분석

이 절에서는 온톨로지 구축을 위한 온톨로지 도메인 레벨을 찾기 위해서 사용한 연관규칙 알고리즘에 대한 실험 예를 소개한다. 이 실험은 웹 사이트를 통해서 가져온 관광 정보들을 XML 문서로 변환 과정을 수행하여 관계형 데이터베이스에 저장된 XML문서의 태그의 수가 2,500개로 이루어진 데이터를 기반으로 실험한다. 데이터베이스에 저장된 훈련데이터 집합 중에서 관광지 정보에 대한 테이블 스키마는 다음과 같다.

〈표 4〉 관광지 정보 테이블 스키마

테이블 명세서						
테이블 ID	관광지 정보					
NO	컬럼 ID	컬럼명	Type	Length	NULL	Key
1	지역	지역이름	varchar2	256		P.K
2	도시	도시이름	varchar2	256		
3	관광지	관광지이름	varchar2	256		
4	숙박시설	숙박시설 이름	varchar2	256		

실험은 적절한 임계 값 설정을 위해 먼저 최소 지지도 변화에 따른 빈발 항목과 실행시간의 변화에 대해서 알아본다. 최소 지지도의 값은 절대 지지도와 상대 지지도로 나누어서 실험을 하였고 다음으로는 최소 신뢰도와 세 가지의 최소 지지도 값에 따라서 각각 생성되는 연관규칙의 개수의 변화과정에 대해서 비교 설명한다.

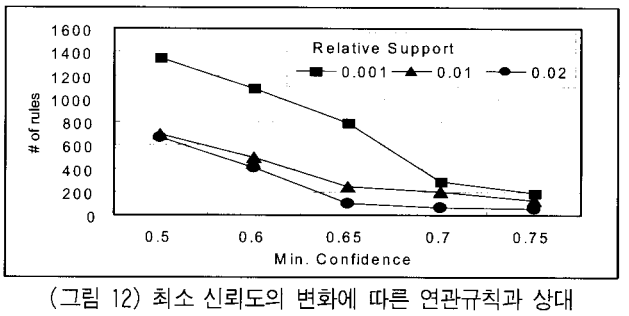
다음 (그림 11)은 각각 최소 절대 지지도와 최소 상대 지지도 값의 변화에 따른 빈발 항목집합의 생성과정과 실행시간의 변화 과정을 보여주고 있다.

이 실험은 최소 신뢰도의 임계 값을 0.5로 정하고, 최소의 절대 지지도 값의 변화에 따라서 빈발 항목 집합의 생성과정과 알고리즘을 수행하는데 걸리는 실행시간의 변화과정을 볼 수 있다. 지지도의 증가에 따라 빈발 항목과 실행 시간이 감소하고 있다. 위의 예를 통해서 연관 규칙에서는 지지도가 작을수록 데이터베이스에서 생성되는 후보 항목 수가 많아지고 그 결과로 수행시간도 오래 걸린다는 것을 확인할 수 있다.

다음은 최소 신뢰도 값의 변화에 따른 규칙 생성과 상대 지지도의 변화 과정에 대해서 알아본다. 상대 지지도를 0.001, 0.01, 0.02로 정해 놓고, 최소 신뢰도 값의 변화에 따른 규칙 생성의 변화과정을 다음 (그림 12)의 그래프를 통해서 확인할 수 있다.

연관규칙의 신뢰도는 조건부와 결과부를 함께 포함하고 있는 비율을 나타내는 값으로 최소 신뢰도 값이 커질수록 연관규칙의 생성은 줄어든다.

XTM을 이용하여 온톨로지를 구축한 기존의 검색 시스템[6,22,23,24]에서는 데이터 마이닝 기법을 적용하고 있지 않다. 그러나 이 논문에서는 온톨로지 구축을 위한 도메인을

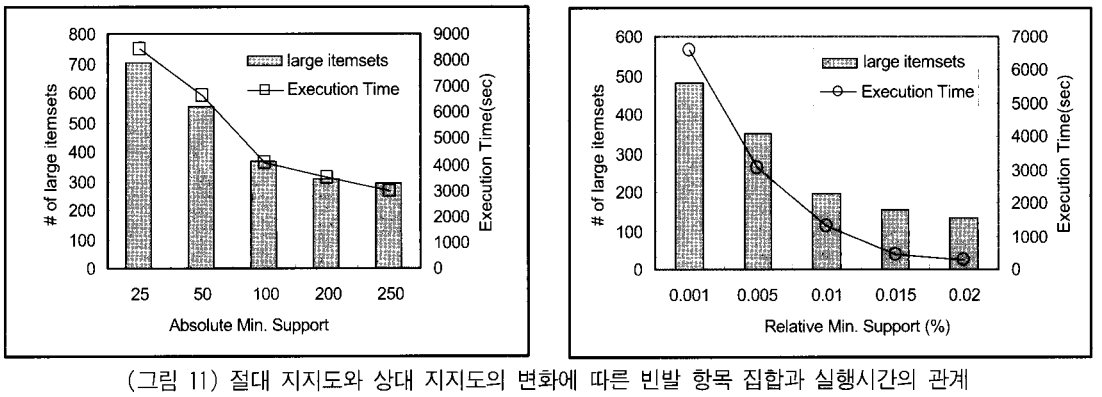


(그림 12) 최소 신뢰도의 변화에 따른 연관규칙과 상대 지지도의 변화

자동으로 결정하기 위해서 웹상의 데이터 중에서 연관성이 있는 데이터들을 대상으로 연관규칙 알고리즘을 적용하였고, 이를 통해 찾아진 결과들을 이용하여 온톨로지를 자동으로 구축하였다. 그러므로 이 논문에서 제안하는 온톨로지 기반의 정보 검색 시스템은 기존의 온톨로지 구축과정에서 도메인 설계에 소요되는 시간을 상당히 줄일 수 있다는 장점이 있다. 또한 사용자의 검색과정에서도 사용자가 원하는 정보를 빠르고 정확하게 제공할 수 있다는 장점을 제공한다. 예를 들어 기존의 검색 시스템은 "제주도에 있는 숙박 시설"을 키워드로 검색할 경우, 키워드인 "제주도"와 관련된 사이트들과 "숙박시설"과 연관 있는 다양한 사이트들을 모두 디스플레이 해준다. 이것은 사용자가 원하는 검색과 더불어 필요하지 않은 정보까지 제공하게 되므로 사용자가 다시 정보를 걸러내야 하는 어려움이 있었다. 그러나 이 논문에서 제안한 시스템은 온톨로지를 기반으로 하는 계층구조의 키워드 메뉴를 이용하며, 사용자가 원하는 정보와 연관된 정보들만을 신속하고 정확하게 제공할 수 있다는 특징이 있다.

6. 결 론

인터넷 기술의 발전으로 현재의 웹은 대량의 정보가 범람하고 있어서 사용자들은 자신이 원하는 정보에 대한 정확한 검색이 어려운 실정이다. 현재의 웹에 의미정보를 추가하여 정보 검색에 효율성을 기하기 위해 새로운 패러다임인 시맨틱 웹이 대두되고 있다. 시맨틱 웹은 온톨로지를 기반으로 인간과 기계가 이해할 수 있는 기법을 프로그램화 시켜서 웹에 적용시키고자 하는 것을 목적으로 한다.



(그림 11) 절대 지지도와 상대 지지도의 변화에 따른 빈발 항목 집합과 실행시간의 관계

현재의 웹은 HTML을 이용하여, 사용자 질의와 관련된 웹 페이지를 중복적으로 디스플레이 해주는 정도의 단순성 때문에 풍부한 정보를 제공해 주지 못하는 단점이 있다. 그래서 현재의 웹 문서뿐만 아니라 공공 문서 등에 XML을 이용하여 표준화 작업이 진행 중이기 때문에 XML문서가 증가하고 있으며, XML은 메타 데이터를 표현해서 풍부한 정보를 전달 해 주는 장점이 있다. 따라서 이 논문에서는 이러한 XML 문서를 기반으로 데이터 마이닝 기법을 이용하여 시맨틱 웹의 기반인 온톨로지를 구축하는 방법을 제안하였다. 온톨로지를 수동으로 구축하는 경우 비용과 시간이 낭비되는 문제가 발생하여 비효율적이므로, 이러한 단점을 해결하기 위해서 온톨로지 구축의 반자동화 방식을 제안하였다. 즉, 특정한 도메인 내의 온톨로지를 구축하기 위해서 어떠한 타입, 그리고 얼마나 많은 개념을 사용할 것인가 하는 문제를 해결하기 위해서 데이터 마이닝을 이용하였다.

XML 문서의 태그에 대해 연관규칙을 적용하여 빈발하게 발생하는 빈발 패턴을 찾아내어 서로 관련 있는 개념의 쌍을 추출하여 온톨로지 자동 생성의 기반을 마련하여 온톨로지를 구축하였다. 이와 같은 빈발 태그를 온톨로지 언어중의 하나인 XML Topic Maps를 이용하여 온톨로지를 구축하였고, 공개 소스인 토픽맵 엔진인 TM4J를 이용하여 온톨로지를 자동 구축한 시맨틱 웹 엔진을 구현하였다. 온톨로지 기반의 시맨틱 웹 정보 검색 기법은, 기존의 웹 시스템에 의미적인 요소를 추가하여 사람과 기계가 이해할 수 있는 있는 방식을 구현하여, 시맨틱 적인 요소를 추가하므로써 사용자의 정보 검색 요구에 정확한 정보를 제공할 수 있다. 그리고 XPointer와 Hytime 기법을 이용하여 웹의 링크를 다양하고 편리하게 한다.

앞으로는 다양한 데이터 마이닝 기법을 적용하여 다양한 분야에 대한 온톨로지의 자동 구축과, RDF(S), OWL, DAML 등의 다양한 온톨로지 언어를 적용하는 연구가 필요하다. 그리고 이 논문에서 제안한 시맨틱 웹을 이용한 정보검색 시스템과 다른 시스템과의 성능비교를 계속 할 것이며, 또한 이 논문에서 제시한 정보검색 시스템의 재현율(recall)과 적합율(precision)에 대해 제시하여 사용자에게 어느 정도의 정확한 정보를 줄 수 있는가에 대한 연구를 계속 할 것이다.

참 고 문 헌

[1] T.R.Gruber, "Toward principles for the design of ontologies used for knowledge sharing", Int. J.Human - Computer Studies, Vol.43, pp.907~928, 1995.

[2] M.Ushold, M.Gruninger, "Ontologies: principles, methods and applications", The Knowledge Engineering Review, Vol.11, No.2, pp.93~136, 1996.

[3] S. Staab, H. P. Schnurr, R. Studer, Y. Sure, "Knowledge processes and ontologies", IEEE Intelligent Systems, Special Issue on Knowledge Management, Vol.16 No.1, pp.26~34, 2001.

[4] Steve Pepper, "The TAO of Topic Maps", XML Conference & Exposition, 2000.

[5] S. Pepper, B. Moore, "XML Topic Maps(XTM) 1.0", TopicMaps.Org.

[6] Koung-lung Lin, Yen-jen Oyang, "Knowlede Management for a Buddhism Digital Archive with Topic Map", ICDAT 2002, pp.91~101, 2002.

[7] D. Braga, A. Campi, S. Ceri, M. Klemettinen, P. Lanzi, "Discovering interesting information in XML data with association rules", SAC, Proceedings of the 2003 ACM symposium on Applied computing table of contents, pp.450~454, 2003.

[8] R. Agrawal, T. Imielinski, A. N. Swami, "Mining association rules between set of items in large database", Proceedings of ACM SIGMOD Conference on Management of Data(SIGMOD '93), pp.207~216, 1993.

[9] R. Agrawl, R. Srikant, "Fast Algorithms for Mining Association Rules", Proceedings of the VLDB, pp.487~499, Santiago de Chile, Chile, September, 1994.

[10] D. Braga, A. Campi, S. Ceri, M. Klemettinen, PL. Lanzi, "A Tool for Extracting XML Association Rules from XML Documents", in Proceedings of IEEE-ICTAI 2002, pp.57~64, Washington DC, USA, November, 2002.

[11] Q. Ding, K. Ricords, J. Lumpkin, "Deriving General Association Rules from XML Data", DBLP:conf/snpd/2003 pp.348~352. 2003.

[12] A. Termier, M-C. Rousset, M. Sebag, "TreeFinder: a Fast Step towards XML Data Mining", In Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM 2002), pp.450~457, 2002.

[13] A. Maedche, S. Staab, "Discovering Conceptual Relations from Text", Technical Report 399, Institute AIFB, Karlsruhe University, 2000.

[14] A. Maedche, S. Staab, "Semi-Automatic Engineering of Ontologies from Text", Proceedings of the 12th International Conference on Software Engineering and Knowledge Engineering, 2000.

[15] R. Srikant, R. Agrawal, "Mining Generalized Association Rules", In Proc. of VLDB '95, pp.407~419, 1995.

[16] http://www.cs.toronto.edu/tox/toxgene/index.html

[17] http://www.cogsci.princeton.edu/~wn/wn2.0

[18] http://www.ontopia.net

[19] http://www.hibernate.org

[20] Jacky W. W. Wan, G. Dobbie, "Mining Association Rules from XML Data using XQuery", ACM International Conference Proceeding, Vol.54, 2004.

[21] 이정원, 방건동, 박세형, 백두권 "온톨로지 기반 설계 문서 관리 시스템 설계 및 구현",한국정보 과학회, 제 28권, 1호, pp.79~81, 2001.

[22] 김정민, 박철만, 정준원, 이한준, 정호영, 민경섭, 김형주, "K-Box : 토픽맵 기반의 온톨로지 관리 시스템", 정보과학회 춘계학술대회, Vol.10, No.1, pp.1~13, 2004.

[23] 김정민, 박철만, 정준원, 이한준, 정호영, 민경섭, 김형주, "온톨로지 기반의 지식맵 서비스 시스템의 설계 및 구현", 한국정보과학회 학술발표논문집 제30권 제1호(A) pp.527~529, 2003.

[24] 정호영, 김정민, 정준원, 김형주, "XTM 기반의 지식맵", 데이터베이스연구회 학회지 Vol.19, No.01, pp.0038~0047, 2003.

[25] 오장근, "유로워드넷 기반의 어휘 데이터베이스 활용을 위한 한국어-독일어 ILI 대응 방법론 연구", 한국독일어문학회 추계 학술대회, 2002.

[26] 박명제, 민준기, 윤정희, 안재용, 정진완, "관계 형 데이터베이스와 XQuery를 이용한 XML 문서의 저장 및 검색 시스템", SIGDB-KISS Vol.18, No.02, 2002.
[27] 장형화, 홍의경, "관계 데이터베이스 시스템 기반의 XQuery 질의 처리기 설계", 정보과학회 추계 학술대회 Vol.30, No.2-2, pp.0106~0108, 2003.
[28] 최규원, 정채영, 김영옥, 김영균, 강현석, 배종민, "관계형 데이터베이스에서 XML 뷰 기반의 질의 처리 모델", 한국정보처리학회 논문지 Vol.10, No.02, pp.0221~0232, 2003.

구 미 숙

e-mail : gumisug@dblab.chungbuk.ac.kr
1986년 충남대학교 영어영문학과(문학사)
2004년 충북대학교 전자계산학과(이학석사)
2004년~현재 충북대학교 전자계산학과 박사과정
관심분야 : XML, 시공간 데이터베이스, 유비쿼터스 컴퓨팅, 바이오 인포매틱스, 데이터 마이닝

황 정 희

e-mail : jhhwang@dblab.chungbuk.ac.kr
1991년 충북대학교 전산통계학과(이학사)
2001년 충북대학교 전자계산학과 (이학석사)
2005년 충북대학교 전자계산학과 (이학박사)
현 재 남서울대학교 컴퓨터학과 전임강사
관심분야 : XML, 데이터 마이닝, 능동 데이터베이스, 유비쿼터스 컴퓨팅, 시공간 데이터베이스

류 근 호

e-mail : khryu@dblab.chungbuk.ac.kr
1976년 숭실대학교 전산학과(이학사)
1980년 연세대학교 전산전공(공학석사)
1988년 연세대학교 전산전공(공학박사)
1976년~1986년 육군군수 지원사 전산실 (ROTC 장교), 한국전자통신연구원 (연구원), 한국방송통신대학교 전산학과 조교수
1989년~1991년 Univ. of Arizona Research Staff (TempIS 연구원, Temporal DB)
1986년~현재 충북대학교 전기전자컴퓨터공학부 교수
관심분야 : 시간 데이터베이스, 시공간 데이터베이스, Temporal GIS 및 지식기반 정보검색 시스템, 데이터 마이닝 및 데이터베이스 보안, 바이오 인포메틱스

홍 장 의

e-mail : jehong@chungbuk.ac.kr
1988년 충북대학교(학사)
1990년 중앙대학교(공학석사)
2001년 한국과학기술원(공학박사)
2002년 국방과학연구소 선임연구원
2003년 국가기술지도(NTRM) 및 국제협력제도 작성위원, 과기부
2002년~2004년 (주) 솔루션링크 기술연구소장
2004년~현재 충북대학교 컴퓨터 공학 조교수
관심분야 : 소프트웨어공학, 임베디드 소프트웨어, 소프트웨어 품질공학, 소프트웨어 프로세스