

개념 망을 이용한 키워드 기반의 효율적인 정보 검색 시스템 설계[☆]

Design of an Efficient Keyword-based Retrieval System Using Concept lattice

마 진¹ 전 인 호¹ 최 영 근^{2*}
Jin Ma In ho Jeon Young keun Choi

요 약

본 논문에서는 개념망을 이용한 효율적인 정보검색을 위한 방법을 제안한다. 본 논문은 일반적인 개념망을 기반으로 시스템을 설계하였기 때문에 온톨로지와 접근 방식은 같지만 사용자가 보다 효율적으로 정보검색을 하고자 하는 객체와 개념사이의 협업 관계를 구축하여 사용할 수 있도록 개념망을 제안한다. 제안한 시스템은 다음과 같다. 첫 번째, 입력 키워드 중심의 키워드 개념망과 전문가 그룹이 추천한 전문가 개념망 그리고 테마 개념망 이러한 세 종류의 개념을 이용하여 협업적 검색을 하며, 이를 기반으로 사용자가 원하는 정보를 검색할 수 있는 효율적인 검색 시스템을 제안한다. 그리고 전문가 개념과 키워드 개념이 결합되어 키워드의 빈도 및 카테고리의 빈도를 제공함으로써, 사용자가 입력한 검색어와 관련된 키워드를 추천하는 역할을 할 수 있다. 그리고 테마 개념망을 이용하여 사용자의 관심 테마에서 사용되는 키워드 또는 카테고리를 알려주는 기능도 제공한다. 두 번째, 사용자가 입력한 키워드가 없을 경우 2차 검색을 통해 입력 키워드와 관련 있는 키워드를 제공해줌으로써 관련키워드를 이용하여 검색의 목적달성이 가능하다. 세 번째, 이러한 정보들은 대부분 분산되어 관리되고 있기 때문에 이렇게 분산되어 관리되는 정보는 표현방식이 다를 뿐만 아니라 시간에 따라 정보가 변하게 된다. 따라서 분산된 정보의 효율적 데이터 접근 및 통합을 위해 XMDR(eXtended Mata-Data Registry)을 이용하였고, 본 논문에서는 분산된 데이터를 통합하기 위한 기법 및 검색 시스템을 제시한다.

☞ 주제어 : 개념망, 검색 시스템, 키워드 기반, 데이터 통합

ABSTRACT

In this thesis was conducted to propose a method for efficient information retrieval using concept lattices. Since this thesis designed a new system based on ordinary concept lattices, it has the same approach method as ontology, but this thesis proposes new concept lattices to be used by establishing collaborative relations between objects and concepts that users are likely to search information more efficiently. The system suggested by this thesis can be summarized as below.

Firstly, this system leads to a collaborative search by using Three kinds of concepts, such as keyword concept lattices, which focus on input key words, expert concept lattices recommended by experts and theme concept lattices, and based on these 3 concept lattices, it will help users search information they want more efficiently. Besides, as the expert concept and the keyword concept become combined, further providing users with the frequency of keyword and the frequency of category, this system can function to recommend key words related to search words entered by users. Another function of this system is to inform users of key words and categories used in users' interested themes by using the theme concept lattices.

Secondly, when there is not keyword entered by a user, it is possible for users to achieve the goal of search through the secondary search when this system provides them with key words related to the input keyword.

Thirdly, since most of the information is managed while being dispersed, such dispersed and managed information not only has different expression methods but changes as time goes. Accordingly, By using XMDR for efficient data access and integration of distributed information, this thesis proposes a new technique and retrieval system to integrate dispersed data.

☞ keyword : Concept lattice(Concept net), Retrieval System, Keyword-based, Data-Integration

1. 서 론

현재 인터넷이 널리 보급됨에 따라 방대한 양의 지식

과 정보가 네트워크를 통해 유기적으로 연결되어있다. 그로 인해 누구나 쉽게 방대한 양의 지식과 정보에 접근할 수 있게 되었다. 하지만 너무 많은 정보량으로 인해

¹ Korea Institute of Science and Technology Information(KISTI), Daejeon, 305-806, Korea

² Dept. of Computer Software, Kwang-Woon University, Seoul, 139-701, Korea

* Corresponding author (ygchoi@kw.ac.kr)

[Received 21 November 2014, Reviewed 24 November 2014, Accepted 30 March 2015]

☆ 이 논문은 2014년도 정부 (미래창조과학부)의 재원으로 한국연구재단 첨단사이언스·교육허브개발사업의 지원을 받아 수행된 연구임(No. NRF-2011-0020576).

사용자가 원하는 의미 있는 정보를 검색하는 것은 오히려 더 어려워지고 있다. 이에 따라 수많은 데이터들 속에서 의미 있는 정보를 검색하기 위한 연구가 활발히 진행되어 왔다. 이러한 연구의 하나로 키워드 검색이 제안되었다[1]. 하지만 키워드 검색은 키워드를 포함한 정보들을 색인하기 때문에 색인할 양이 현저히 많아진다. 그에 따라 효율적 인덱스 처리에 대한 연구도 진행되었다[2].

한편 키워드 검색에서는 사용자의 질의어와 일치하는 인덱스를 가지고 있는 문서를 검색하기 때문에 인덱스의 정보가 부족한 경우 검색의 결과는 너무 많거나 또는 의미가 없는 경우가 많다. 이러한 문제를 해결하기 위해 각종 서비스에서는 의미 정보들을 잘 구축해놓은 시소러스, 언어자원, 지식베이스 등의 유의어, 동의어, 관련어 정보 등을 활용하여 연관 검색 결과를 얻을 수 있는 의미 기반 서비스를 하는 연구가 활발히 진행되고 있다[3]. 현재 일반적인 검색 사이트에서는 사용자가 요구하는 적합한 용어를 찾기 위해 정규화 된 단어들을 바탕으로 벡터 공간 모델[4]이나 확률 모델[5]을 이용하여 색인 하거나 빈도수에 따라 색인화를 한다. 하지만 이러한 검색 방법은 트리구조의 색인에만 의존하고 있어 다른 정보와의 연관관계에 대한 검색이 어렵다. 따라서 개념 기반의 정보검색은 초기 검색어에서 개념 확장을 수행한 후 확장된 검색어를 수행한다. 이렇게 확장된 키워드 검색 기반의 시스템에서 사용자는 자신이 원하는 자료를 검색할 수 있도록 할 수 있다. 한편, 키워드 검색의 하나로 개념 망을 이용한 키워드 정보 검색 시스템 또한 제안되었다.

본 논문에서는 기존의 키워드 검색이 가지는 문제점을 해결하기 위해 개념 망 기반 검색 모델 및 시스템을 제시한다. 이 시스템은 키워드를 기반으로 정보를 효율적으로 검색을 할 수 있도록 개념 망을 협업적으로 사용한다. 구체적으로, 본 논문에서 제안하는 기법은 키워드 검색을 통한 키워드 개념 망과 전문가 개념 망 그리고 테마별 개념 망을 사용한다. 이후 구축한 전문가 개념 망을 통해 키워드 개념 망과 협업적 검색을 수행하고, 이를 기반으로 사용자가 원하는 정보를 효율적으로 검색할 수 있도록 한다. 사용자 검색 키워드에 대한 키워드 개념 망에서 스키마 이질성을 해결하기 위해 XMDR(eXtended Meta-Data Registry)[13]을 이용하였다. 본 논문의 구성은 다음과 같다. 2장은 관련 연구로 제반 기술에 대해서 서술하고, 3장은 개념 망의 구성과 설계에 대해 기술하고, 4장은 시스템을 구성하는 구성요소와 각 구성요소의 역할과 시스템 운용에 대해 기술하고, 5

장은 적용 및 평가를 수행한다. 그리고 6장은 결론과 향후 연구에 대해 기술한다.

2. 관련 연구

본 절에서는 개념 망을 다른 정보 검색 관련 연구와 비교 분석한다. 먼저 색인(indexing)은 어떤 문서를 다른 문서들로부터 구별할 수 있는 단어 또는 단어 구를 추출하여 정보 검색에 활용한 것으로 현재 대다수의 검색엔진이 사용하고 있는 기술이다. 색인의 경우, 사용자가 검색을 위해 선택한 검색어에 따라 쓸모없는 검색 결과가 나올 수도 있다. 또한, 사용자가 원하는 적절한 결과를 얻기 위해 일반적으로 수차례의 피드백 과정이 필요하다. 반면에, 개념 망은 사용자가 입력한 검색어와 관련한 개념들을 미리 제시해 사용자가 필요에 따라 새로운 검색어를 취하거나 또는 추가, 삭제 하도록 함으로써, 검색의 효과를 향상시킬 수 있어 정보 검색의 프리프로세서 역할을 한다.

2.1. FCA 개념 망

개념 망은 주어진 문서에서 제시된 개념을 추출하고 그 추출된 개념들 사이의 연관성을 추출하여 관계성이 높은 개념 사이에서의 네트워크를 구성한다. 이런 개념 망은 FCA로 표현할 수 있다. FCA(Formal Concept Analysis)는 특정영역의 지식이나 자료를 모델링하기 위한 방법으로 자료 집합 사이의 개념적 구조를 조직화하기 위하여 수학적 사고로 접근한 자료 분석의 한 이론이며, 1982년 독일의 한 수학자에 의해 개발되었다[6]. FCA는 특정 도메인에 존재하는 객체들과 속성들 사이의 이항관계로부터 개념을 추출하여 개념들 사이의 관계를 망(lattice) 구조로 계층화 시킨다. 개념은 객체와 속성사이의 관계를 반영하며 하나의 개념은 객체집합과 그 객체 집합의 원소들과 관계가 있는 속성집합의 쌍으로 이루어진다. 개념망은 개념들 사이에 존재하는 상·하위(subconcept superconcept) 개념 관계를 망구조로 표현한 계층적 개념적 클러스터링이며, 객체들과 속성들 사이의 모든 관계를 표현한 의미 망(semantic net)으로도 볼 수 있다. FCA는 자료 분석을 위한 여러 응용 분야에 적용되었으며 많은 연구들이 FCA의 개념 망을 정보 검색에 적용 발전시켰다[7, 8, 9, 10]. FCA는 세 개의 기본 구성요소 ; 배경도(formal context), 개념(formal concept), 개념 망(concept lattice)으로 이루어진다[11, 12].

2.2. XMDR(eXtended Meta-Data Registry)

XMDR(eXtended Meta-Data Registry)은 XML기반의 관계형 데이터베이스 메타데이터를 객체지향 데이터베이스에 저장하는 기술과 그리드상의 분산된 데이터간의 이질성을 해결하는 MDR(Meta-Data Registry), 그리고 데이터의 효율적인 이용을 위한 연관성을 정의한 온톨로지 시소러스(Ontology thesaurus)를 결합한 것이다[13].

XMDR은 ISO/IEC 11179[14]에서 제안한 정보 공유 및 교환을 위한 표준으로 현재 많은 프로젝트가 진행 중에 있다[15]. 특히, ISO/IEC 11179-3에서 공유 데이터의 관리를 위한 메타 모델, 기본 속성이 제시되었는데, 메타 모델은 의미적인 내용과 그리드 환경하의 사용자들이나 정보처리 시스템간의 공유되는 데이터 요소의 구문을 위한 표준과 안내를 제공하고 있다. XMDR은 MDR을 보다 더 확장한 개념으로 이전의 MDR에 전문 용어, 온톨로지 등의 기능을 확장한 것이다. XMDR의 각 구성 요소들은 ISO/IEC 11179-3에서 제안한 데이터의 속성 명세를 따르며 메타 시맨틱 온톨로지(MSO:Meta Semantic Ontology)와 메타 로케이션(ML:Meta Location), 인스턴스 시맨틱 온톨로지(InSO:Instance Semantic Ontology)로 구성되어 있다[16, 17]. MSO는 메타데이터 간의 관계성과 이질성을 해결한다. MSO는 각 레거시 시스템을 통합하기 위한 표준을 결정하고, 이 표준과 관련된 스키마 정보를 통해 스키마 정보 온톨로지를 생성하고, 이것을 통해 메타데이터의 관계성과 이질성을 해결한다. InSO는 실제 데이터 값 사이의 이질성을 해결한다. 실제 데이터는 표현의 차이로 인해 같은 의미를 지녔지만 서로 다른 데이터로 인식 될 수 있다. 이러한 데이터들의 연관성과 의미를 파악하여 온톨로지로 생성하였다. 따라서 본 논문에서는 분산된 환경에서 효율적인 데이터 검색을 위해 XMDR을 이용한다.

2.3. 시맨틱 웹 (Semantic Web)

시맨틱 웹은 컴퓨터가 웹상의 자원(Resource)을 의미 기반으로 표현하고 처리할 수 있도록 하는 기술로서, 기하급수적으로 증가하고 있는 웹상의 자원을 능동적으로 검색할 수 있도록 기반구조를 제공하는 웹 기술이다[18, 19]. 시맨틱 웹의 가장 기본이 되는 URI(Uniform Resource Identifier)는 웹 자원의 고유 식별자로 사용되며, XML namespace 및 XMLSchema는 리소스의 표현을 구조화 할 수 있도록 지원함으로써 문서의 구조적인 측면에서 상호

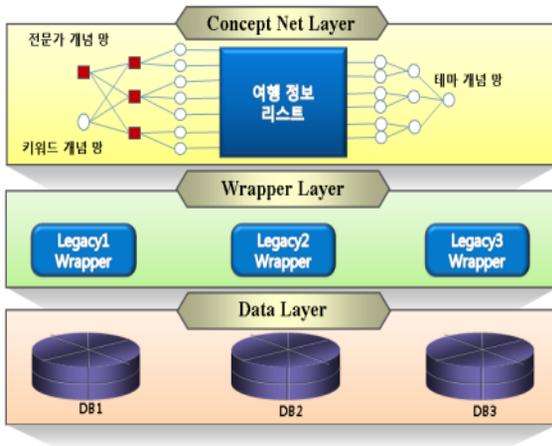
운용성을 제공한다[19].

RDF(Resource Description FrameWork)와 RDF Schema는 웹상의 자원(Resource)들을 정의하고 기술하기 위한 XML 기반의 데이터 모델로서, 컴퓨터가 리소스를 의미적으로 처리할 수 있는 기반을 제공한다. Ontology는 웹 환경에서 자원에 대한 개념과 관계를 정의함으로써 웹상의 자원에 대한 의미를 정의한다. 이렇게 기술된 웹상의 자원들은 Logic, Proof, Trust 계층의 요소기술들을 통해 자원에 대한 지능화된 접근과 신뢰성을 제공받게 된다[19].

특히, RDF와 Ontology는 시맨틱 웹의 핵심적인 기반 구조로서의 역할을 담당한다. RDF는 데이터를 정의하기 위해 자원간의 관계를 Triple구조(subject-predicate-object)로 표현함으로써 컴퓨터가 의미를 처리할 수 있는 기반을 제공한다[20]. 온톨로지(Ontology)[18]는 웹 환경에 존재하는 모든 자원을 개념(Concept 또는 Class)으로 정의하고 그 의미적인 처리를 위해서 그 개념간의 관계를 속성(Property)으로 정의함으로써, 컴퓨터가 웹 자원들을 의미적으로 이해하고 처리할 수 있는 기반을 제공한다. 온톨로지를 정의하기 위한 여러 가지 방법이 있지만, 특히 OWL(Web Ontology Language)[21]은 W3C에서 웹 자원의 관계를 풍부하게 정의할 수 있도록 제공하는 온톨로지를 기술하기 위한 언어이다.

3. 개념망을 이용한 정보검색 시스템 설계

현재까지 많은 검색 방법이 제안되었지만 이전 연구들은 한 측면에서 초점을 맞춘 정보추출 방법이었다. 그러나 본 연구에서는 관련 연구에서 살펴 본 개념 망을 이용하여 검색 및 추천을 위한 통합된 정보추출 모델을 제안한다. 본 논문의 개념 망 모델은 키워드 유사성과 시소러스를 해결하기 위해 전문가 개념 망을 Knowledge Base에 구축한다. 그리고 사용자의 검색을 기반으로 생성되는 키워드 개념 망과 전문가 개념 망 그리고 분야 망을 결합한 개념 망을 이용하여 사용자에게 검색의 결과를 나타낼 수 있다. 본 논문에서 제안하는 개념 망 모델은 그림 3.1과 같이 3계층으로 구성된다. 개념 망 관리를 위한 개념 망 계층(Concept Net Layer), 데이터 통합을 위한 허브 역할을 수행하는 데이터 허브 계층인 데이터 래퍼 레이어(Data Wrapper Layer) 그리고 실제 데이터에 해당하는 데이터 레이어(Data Layer)가 있다.



(그림 3.1) (개념 망 모델)

(Figure 3.1) (Concept lattice Model)

- **Concept Net Layer** : 개념 망 계층은 키워드 개념 망과 테마 개념 망, 전문가 개념 망을 결합하여 생성된 개념 간의 네트워크이다. 키워드 개념망은 사용자가 입력한 키워드를 기반으로 전문 정보를 검색하고, 검색된 전문 정보의 키워드를 추출하여 키워드 간의 개념 관계를 분석한다. 이 개념 관계는 키워드 간의 관계를 쌍으로 표현할 수 있다. 쌍으로 표현된 개념 관계를 결합하여 키워드 개념 망을 생성한다. 이를 바탕으로 미리 생성된 테마 개념망과 전문가 개념 망을 결합한 구조를 생성하고 관리하는 계층이다.
- **Wrapper Layer** : 개념 망을 생성하기 위한 기본적인 데이터 소스의 접근이나 정보 검색을 위한 데이터 접근은 통일된 방식으로 진행되어야 한다. 이를 위해서 분산된 데이터 소스의 접근은 이질적 문제를 해결이 우선되어야 한다. 이를 해결하기 위한 MDR 기술을 바탕으로 데이터들의 구조적 이질성을 해결할 수 있도록 한다. MDR은 분산 환경의 데이터를 접근할 수 있도록 표준항목을 만들고, 이 표준항목 바탕의 질의로 변환하여 로컬 시스템에 접근하여 전문 정보를 추출할 수 있도록 수행하는 계층이다.
- **Data Layer** : 실제 데이터 소스들이 저장되어 있는 데이터 저장소 계층으로 정보 검색을 위한 원천이다.

3.1. 개념 망 모델

본 절에서는 그림 3.1의 개념 망 모델 중 Concept Net Layer에 해당하는 3가지 개념 망에 대하여 기술한다.

3.1.1 개념망 구성요소

개념 망은 특정 도메인을 포함되는 주요 개념과 이들 개념 사이의 관계성을 망으로 표현 한 것이다 이러한 개념 망은 정보 검색을 위한 프리프로세서로 활용할 수 있다. 따라서 일반적인 키워드 검색 보다 정보를 효율적으로 검색할 수 있다. 본 논문에서는 구성하고자 하는 개념 망은 세 개의 개념 망으로 구성되며 키워드 개념 망, 전문가 개념 망, 테마 개념 망으로 구성된다.

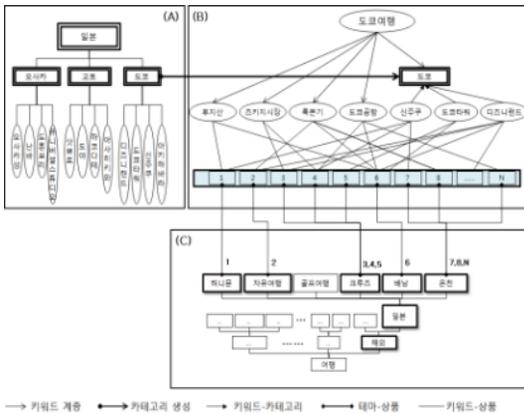
- **키워드 개념 망** : 키워드 개념망은 사용자가 입력한 키워드를 기준으로 해당 키워드를 포함하는 ITEM을 검색하여 검색된 결과에 대한 키워드를 추출하고 키워드간의 빈도를 통계적 방법으로 분석한다. 관련 키워드를 추출하기 위해서는 사용자가 입력한 키워드와 같이 포함된 키워드들을 분석하여 사용자가 입력한 키워드와 관련 키워드를 추출한다. 관련 키워드는 아래의 식 1과 같이 구할 수 있다. 관련 키워드는 사용자가 입력한 키워드를 가진 ITEM을 추출하여 추출된 ITEM에 있는 모든 키워드들이 나타나는 횟수를 분석하여 각 키워드의 분포를 산출하고, 이 수치를 이용하여 전체에 대한 빈도를 구한다. 이 빈도를 기준으로 정렬한 관련 키워드를 이용하여 초기검색에 이용할 수 있다. 그리고 사용자 인터페이스에 출력하는 키워드 리스트는 최소 3회 이상 검색된 키워드 값을 나타내었다.

$$\text{연관키워드의 초기 빈도} = \frac{kf(s)}{\sum_{i=1}^n kf(n)} \quad (1)$$

- **전문가 개념 망** : 카테고리는 여행 정보 구분에 따라 국내와 해외 그리고 해외의 각 나라별 도시에 대하여 도시명과 관광명소가 정의 되어있다. 카테고리에서 제시된 용어는 레거시 시스템에서 사용되는 용어들에 대한 표준으로 각 시스템에서 공유되고 사용자에게 제시되는 용어들의 집합이다. 이와 같이 시스템과 사용자들이 공유할 수 있는 용어들을 제시함으로써 통합-관리 될 수 있고, 이러한

표준은 여행정보 테마 분류와 같은 분류 기준으로 적용시켜 표준을 따라 정의된 용어들을 확장 가능하게 한다. 그리고 카테고리과 키워드 분류 체계는 전문가들의 분류체계를 따른다.

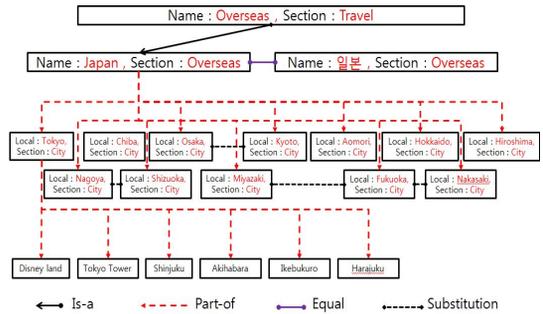
- **테마 개념 망:** 제안 시스템에서 사용하는 데이터는 크게 여행 사이트에 저장되어 있는 대분류와 소분류로 나눌 수 있으며 제안 시스템의 테마 개념망의 분류 체계는 여행 정보 사이트들의 분류 체계에 따라 설계하였다.



(그림 3.2) 개념망의 관계 표현

(Figure 3.2) Representation of the concept lattice of the relationship

그림 3.2는 개념망의 관계를 표현한 것으로 A는 전문가 개념 망을 나타내고 B는 키워드 개념 망, C는 테마 개념 망에 해당한다. 제안 시스템에서 사용하는 데이터는 크게 Knowledge Base DB에 저장되어 있는 전문가 개념 망과 사용자 검색어에 따른 키워드 개념 망으로 나눌 수 있다. ‘도쿄여행’이라는 키워드에 대하여 Knowledge Base의 전문가 개념 망에서는 도시 명을 카테고리로 분류하였으며 카테고리 안에 속하는 키워드(관광명소)는 타원으로 나타 낼 수 있다. 제안 시스템의 전문가 개념 망에서 키워드 분류 체계는 여행정보 사이트들의 분류체계를 따랐다. 사용자가 입력한 키워드에 대한 키워드 개념망은 B에 해당하며 사용자 입력 키워드에 대한 관련 키워드를 나타낼 수 있다. 키워드 개념 망B의 키워드 리스트들과 전문가 개념 망A의 리스트를 비교하여 일치하는 키워드가 있을 경우 그림 3.2와 같이 카테고리를 생성한다. 이 때, 카테고리의 빈도율도 계산을 할 수 있다.



(그림 3.3) 여행 정보 전문가 개념 망
(Figure 3.3) Travel Experts Concept lattice

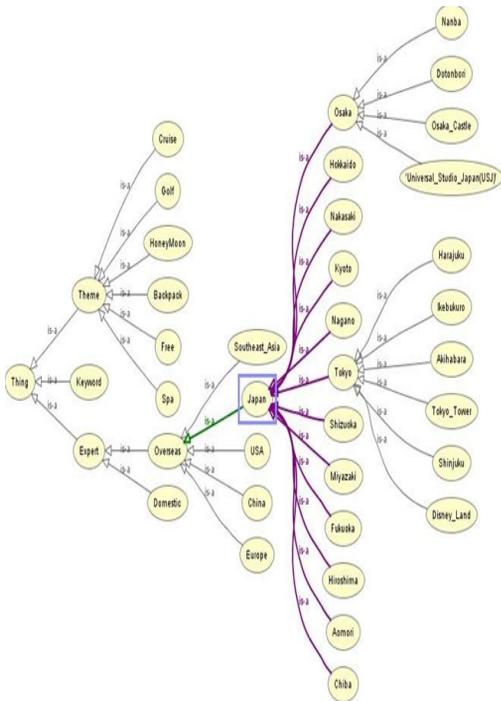
$$category\ frequency = \sum_{i=1}^n kf(n) \quad (2)$$

식 (2)의 카테고리 빈도(category frequency)는 전문가 개념트리에 의해 생성된 카테고리의 빈도를 나타내며, 카테고리에 속하는 관련 키워드들의 합이다. 도쿄 카테고리에 속하는 키워드인 디즈니랜드, 도쿄타워, 신주쿠이 3가지 키워드의 빈도를 합이 카테고리 빈도의 값이 된다. 그림 3.2에서는 카테고리 빈도가 도쿄 카테고리에서 사용되었다. 그리고 효율적인 검색을 위하여 이 카테고리 빈도 값을 사용자에게 알려주도록 하였다.

그림 3.3은 그림 3.2의 (A)에서 표현했던 지식베이스에 구축된 전문가 개념 망을 상세하게 나타낸 것이며 사용자 검색 키워드와 그에 따른 키워드 개념 망과 연계된다. 그림 3.3의 전문가 개념망은 여행 정보 사이트의 해외여행정보에서 일본에 대한 정보를 나타낸 것으로 일본의 지역정보와 도시정보 그리고 각 도시의 관광명소를 개념 망으로 나타냈으며 사용자 검색에 따른 키워드 개념망의 키워드 값과 전문가 개념망의 키워드 값의 일치를 확인 함으로써 사용자에게 효과적으로 정보를 제공할 수 있다. 그리고 전문가 개념 망에서 표현되는 개념 관계는 동등관계(Equal), 상속관계(Is-a), 부분관계(Part-of), 대체관계(Substitution)으로 표현하고 각 관계는 다음과 같이 정의된다.

- 동등관계(Equal) : 표현은 다르지만 의미상 동등한 의미를 가지는 관계를 표현 하는 것으로 InSO의 정의에 따른다. 그림 3.3에서 일본과 Japan은 동등 관계이다.

- **상속관계(Is-a)** : 분류 계층상 상속된 하위 개념과 관계를 표현한다. 그림 3.3의 여행 정보 기반의 전문가 개념 망에서 일본은 해외여행에 속하는 상속 관계이다.
- **부분관계(Part-of)** : 관계상 하위개념이 상위개념의 일부가 되는 관계를 표현한다. 그림 3.3의 여행 정보 기반의 전문가 개념 망에서 시(City)는 일본에 포함되는 관계이다.
- **대체관계(Substitution)** : 대체되어 사용될 수 있는 관계를 표현한 것으로, InSO의 정의에 따른다. 예를 들어 일본의 '교토 여행'을 검색할 때 교토(Kyoto)의 인접지역인 오사카를 InSO에서 대체관계로 설정하였다. 만약 키워드 개념 망에 교토에 대한 여행 정보가 없어서 검색을 할 수 없는 상태이고, 오사카(Osaka) 대한 여행정보만 찾을 수 있다면 오사카는 교토의 대체관계로써 오사카의 여행정보를 나타낼 수 있다.



(그림 3.4) Protege를 이용한 전문가 개념 망 설계
(Figure 3.4) Using Protege Expert Concept lattice design

```

<!--
<!-- http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Akihabara -->
<owl:Class rdf:about="http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Akihabara">
<rdf:type rdfs:subClassOf rdfs:resource="http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Tokyo"/>
</owl:Class>
<!-- http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Amori -->
<owl:Class rdf:about="http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Amori">
<rdf:type rdfs:subClassOf rdfs:resource="http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Japan"/>
</owl:Class>
<!-- http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Backpack -->
<owl:Class rdf:about="http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Backpack">
<rdf:type rdfs:subClassOf rdfs:resource="http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Class">
</owl:Class>
<!-- http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Chiba -->
<owl:Class rdf:about="http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Chiba">
<rdf:type rdfs:subClassOf rdfs:resource="http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Japan"/>
</owl:Class>
<!-- http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#China -->
<owl:Class rdf:about="http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#China">
<rdf:type rdfs:subClassOf rdfs:resource="http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Overseas"/>
</owl:Class>
<!-- http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Cruise -->
<owl:Class rdf:about="http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Cruise">
<rdf:type rdfs:subClassOf rdfs:resource="http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Theme"/>
</owl:Class>
<!-- http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Disney_Land -->
<owl:Class rdf:about="http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Disney_Land">
<rdf:type rdfs:subClassOf rdfs:resource="http://www.semanticweb.org/ontology/2011/11/Ontology1323069217774.owl#Tokyo"/>
</owl:Class>

```

(그림 3.5) 전문가 개념망의 XML 문서
(Figure 3.5) Expert Concept lattice of XML document

그림 3.4는 그림 3.3에서 설계한 전문가 개념 망의 내용을 온톨로지 툴인 Protege를 사용하여 구축한 화면이고, 그림 3.5는 Protege를 사용하여 구축한 전문가 개념망의 내용을 XML 형식으로 추출한 화면이다.



(그림 3.6) 여행 테마 개념 망
(Figure 3.6) Travel theme Concept lattice

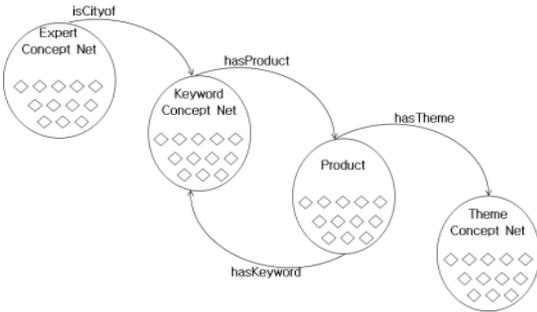
그림 3.6의 분류 개념망은 크게 여행 사이트에 저장되어 있는 대분류 - 중분류 - 소분류로 나눌 수 있으며 제한 시스템의 분류 개념망의 분류 체계는 여행 사이트들의 분류 체계에 따라 설계할 수 있다.

3.1.2 개념 망 사이의 관계

그림 3.7은 3가지 개념 망과 상품간의 관계와 속성을 나타낸 것으로 전문가 개념 망과 키워드 개념 망 사이의 관계와 키워드 개념 망과 상품의 관계, 상품과 테마 개념 망 사이의 관계를 나타내었다. 사용된 속성명은 isCityof, hasProduct, hasTheme, hasKeyword 이다.

isCityof는 사용자 입력 키워드를 기반으로 생성되는 키워드 개념 망과 Protege를 사용해 미리 구축해놓은 전문가 개념 망을 비교하여 사용자가 입력한 키워드의 관

광명소 or 도시명이 전문가 개념 망에 존재하는지를 살펴보는 속성이다. hasProduct는 사용자 입력 키워드가 갖는 상품을 확인하는 속성이고, hasTheme는 어떤 상품이 테마 분류기준에 따라 무슨 테마로 분류되어 있는지를 살펴볼 수 있는 속성이다. 그리고 hasKeyword는 각각의 상품에 사용자 검색 키워드들이 포함되어 있는지를 확인하는 속성으로 hasProduct와 한 쌍을 이루는 속성이다.



(그림 3.7) 개념 망과 상품 간의 관계

(Figure 3.7) The Concept lattice and relationship between Products

3.1.3 XMDR 구성요소

다음은 개념 망을 위한 메타데이터 설계를 위해 XMDR의 정의 및 설계를 기술한 것이다.

- **Meta Semantic Ontology(MSO)** : MSO는 로컬 데이터베이스들의 메타데이터 스키마 정보를 시소러스화 한 것으로, 표준으로 지정한 메타데이터 스키마에 매핑 하여 메타데이터간의 관계성과 이질적인 충돌 문제를 해결할 목적으로 정의한 것이다. 또한 MSO는 스키마 표준인 글로벌 메타데이터 스키마를 로컬 메타데이터 스키마로 변환하기 위해서 매핑 할 때 필요한 정보이다.
- **Meta Location(ML)** : ML은 MSO와 연계하여 로컬 데이터베이스들의 물리적인 위치, 접근권한 정보 등을 등록한 것이다. 이는 데이터 접근 및 통합에서의 상호운용상 필요한 데이터 이주 및 트랜잭션 과정을 비즈니스 프로세스 메시지가 해당 위치에 전달될 수 있도록 정의한 것이다.
- **Instance Semantic Ontology(InSO)** : InSO는 실제 인스턴스 값(value)간의 연관성(association) 정보를 매

핑 구성하여 시소러스화 한 것으로, 인스턴스 값 사이에 의미성, 유사성, 유효성을 고려하여 정의한 것이다. 예를 들어 단위 형식(ex: mm->cm, kg->pound, mm/dd/yy->yy-mm-dd)의 불일치를 충돌 정보로 분류하여 정의한 것이다.

- **MetaData Registry(MDR)** : MDR은 각 로컬 데이터베이스의 메타데이터 개체 스키마를 등록하여 관리하는 것이다. 이 MDR은 글로벌 스키마 영역과 로컬 스키마 영역으로 구성된다.
 - **Global Schema** : 글로벌 스키마는 각 로컬 스키마들의 표준 스키마를 선정하고, 이를 비즈니스 협업에 맞게 구성한 것이다.
 - **Local Schema** : 로컬 스키마는 협업에 참여하는 로컬 데이터베이스의 스키마를 등록한 것이다.

(표 3.2) 인스턴스 값들의 연관성

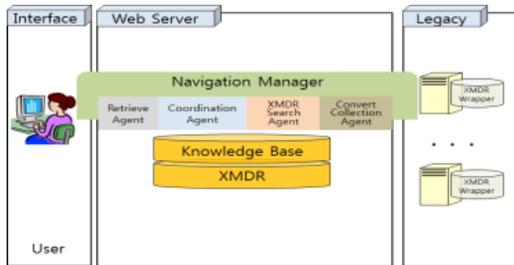
(Table 3.2) Association of Instance Values

연관 유형	설명	사례
equal-association (동등 관계)	값을 표현하는 방식은 다르지만, 의미상 동일한 의미를 가지는 관계.	Tokyo = 도쿄 = 동경
contain-association (포함 관계)	값 사이에 의미적으로 포함적인 관계.	Overseas \ni Japan, China
part-of-association (부분 관계)	값 사이에 의미적으로 상위, 하위 개념관계	Japan > Tokyo

표 3.2는 인스턴스 값 즉, 실제 값 사이에 연관적인 사항을 구조화하여 구체적으로 정의한 것으로 단순한 검색이 아니라 검색 키워드에 대한 의미성, 유사성 등을 고려하여 검색에 대한 결과를 확장하려는 목적이 있다. 예를 들어, 그림 3.3에서 '도쿄'에 대해서 요청한다면 'Tokyo'와 같이 같은 의미를 갖는다. 하지만 'Tokyo'는 다른 관점에서는 'Japan', 'Kanto'에서 상위, 하위 개념으로 표현될 수도 있다. 이런 의미상으로 표현이 다른 것도 같이 검색이 되어 '도쿄'에 정보를 더욱 확대하여 검색할 수 있도록 한다. 본 논문에서의 사용되는 XMDR은 데이터 기본 속성 명세 사항, 메타데이터 스키마간의 이질성 분류, 표 3.2의 인스턴스 값들이 연관성을 고려하여 정의 및 구성된 것으로 효율적으로 데이터 접근 및 통합하는데 용이하다. 이는 XMDR의 구성요소인 MSO, InSO를 정의하는데 참조된다.

3.2. 정보 검색 시스템

본 논문에서 구축하고자 하는 시스템은 그림 3.8과 같이 4부분으로 구성된다. 첫째, 인터페이스는 사용자와 관리자에게 제공할 화면을 구성하고, 구성에 필요한 요소를 Web Server 검색을 통해 인터페이스를 생성하는 역할을 한다. 둘째, Web Server는 데이터의 관계성을 밝혀 관련된 정보를 검색할 수 있도록 제공하고, 각 레거시 시스템의 데이터베이스에 구성된 데이터베이스의 구조나 의미에 대한 관계성을 정의하고, 이를 유지, 관리하는 XMDR로 구성된다. 셋째, 실질적인 검색과 검색된 데이터를 인터페이스에 표현하기 위해 에이전트로 구성되는 코디네이션 관리자 (Coordination Manager) 그리고 넷째, 레거시시스템에 제공된 XMDR 래퍼로 구성된다.

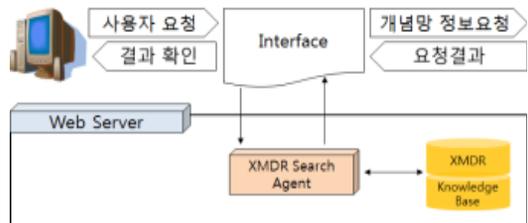


(그림 3.8) (시스템 구성도)
(Figure 3.8) (Configuration of System)

3.2.1 인터페이스(Interface)

인터페이스는 지식베이스와 데이터베이스 표현의 표준에 의해서 생성된다. 지식 정보는 지식베이스에 저장된 전문가 개념 정보를 통해 검색 키워드의 관련정보를 제공받고, 데이터베이스 표준은 XMDR을 검색하여 생성된다. 인터페이스는 코디네이션 관리자의 XMDR 검색 에이전트에 의해 지식베이스와 XMDR을 검색하여 생성한다. 생성된 인터페이스에 의한 질의 내용은 변환 에이전트에 의해 XML_Query로 변환되고, 수행된 결과인 XML 문서를 XSLT변환을 통해 사용자에게 제공한다. 입력 화면 생성, 검색 조건 전송, 검색된 결과를 사용자에게 제공하는 출력 화면 생성이 인터페이스의 역할이며 그림 3.9는 인터페이스의 작업을 개괄적으로 표현한 것이고, 알고리즘 3.1은 인터페이스를 생성하는 알고리즘으로 지식베이스와 XMDR의 검색이 주된 기능이다. 알고리즘 3.1은 사용자 검색요청에 따라 지식베이스의 전문가 개념트리의 카테고리 id와 이름을 검색한 다음 XMDR

검색을 통해 사용자 입력 키워드, 테마, 상품명을 검색하는 과정을 나타낸 것이다. 사용자가 접근하면 웹서버에 접속하여 전문가 개념망의 카테고리를 사용자에게 제공하고, 선택된 키워드를 지식베이스를 통해 확인하여 용어에 대한 관계성을 사용자에게 제공함으로써 정확한 검색이 되도록 한다. 선택한 정보는 XMDR 검색 에이전트에 의해 XMDR에서 표준 항목을 추출하여 입력화면을 생성한다. 인터페이스는 전문가 개념망의 카테고리 정보, 관련 키워드에 대한 정보, 검색의 직접적인 조건이 되는 사용자 요청 정보, 테마 조건으로 구성된다.



(그림 3.9) 인터페이스 생성을 위한 과정
(Figure 3.9) Interface for the Creation Process

```
Void Interface_Agent() {
//지식베이스 검색
kbsearch(cate_id, cate_name);
//XMDR 검색
xmdrsearch(input_keyword, travel_theme, product_title);
}
```

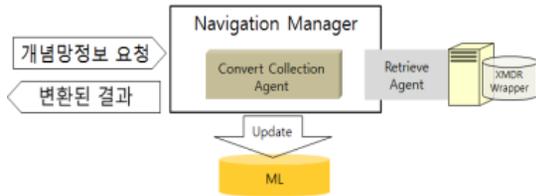
(알고리즘 3.1) (인터페이스 에이전트 알고리즘)
(Algorithm 3.1) (Interface Agent Algorithm)

3.2.2 웹 서버(Web Server)

웹 서버는 협업을 위한 레거시 시스템들의 통합을 위해 3장 에서 소개한 것과 같이 XMDR 계층으로 구성되고, 이 계층은 지식 베이스와 MDR로 구성된 저장소와 데이터의 교환을 위한 에이전트를 관리하는 네비게이션 관리자로 구성된다. 네비게이션 관리자는 XMDR을 검색하는 XMDR 검색에이전트, 추출된 질의정보를 XML_Query로 변환하고 레거시 시스템의 검색 결과인 XML문서를 XSLT를 이용한 출력형식으로 변환하고 결과를 수집하는 변환-수집 에이전트, XML_Query를 레거시 시스템에 전송하여 XMDR래퍼를 통해 질의변경을 발생하도록 하는 검색에이전트 그리고 검색된 결과를 통계적으로 계산하여 빈도를 나타내주는 조정 에이전트로 구성된다.

• 네비게이션 관리자(Navigation Manager)

네비게이션 관리자는 논문에 대한 검색을 수행하기 위한 표준 XML_Query를 생성하고 이를 생성된 검색 에이전트에 의해 논문사이트의 매핑 에이전트로 전송하는 역할을 하고 검색된 XML결과를 XSLT를 이용하여 테이블 형식으로 변환하며 세부적인 내용은 다음에서 다룬다.



(그림 3.10) (네비게이션 관리자)
(Figure 3.10) (Navigation Manager)

㉠ XMDR 검색 에이전트(XMDR Search Agent)

인터페이스를 생성하기 위한 항목들을 가져오기 위해 XMDR에 접근하여 필요한 항목과 개념 망을 이용한 검색을 하기 위해 전문가 개념망의 카테고리 항목을 가져오는 역할을 수행하는 에이전트이다. 인터페이스를 생성하기 위한 에이전트로서, 사용자의 검색요구를 입력받고 XMDR에 접근하여 전문가 개념 망과의 연관 관계와 표준 항목을 검색하는 역할을 한다. 그림 3.10에서 XMDR 검색 에이전트의 역할에 대해 볼 수 있고, 시스템에서 적용되는 알고리즘은 알고리즘 3.2와 같다.

```

Void KnowledgeBase(cate_id, cate_name){
// 선택된 지식 베이스의 정보 검색
Knowledge_Stack = FindKnowledge(Ex_id, Ex_value)
}
void xmdrsearch(MSOID, LOCID){
//MSO와 ML정보를 검색하여 레거시 시스템의 정보 검색
XMDR_Select = search XMDR_LOC information
for (i = 0; i = Last XMDR_Stack; i++)
{
    XMDR_Stack_Legacy_id = XMDR_Legacy_id
}
XMDR Stack = FindXMDR(MSOID, LOCID)
}
    
```

(알고리즘 3.2) (XMDR 검색 에이전트 알고리즘)
(Algorithm 3.2) (XMDR Search Agent Algorithms)

㉡ 변환 수집 에이전트 (Convert Collection Agent)

변환 수집 에이전트는 질의 내용을 XML_Query로 변환하여 레거시 시스템의 데이터베이스로 전송되어 수행하도록 하고, 반환된 결과를 XSLT 변환을 통해 사용자에게

제공하는 변환기능과 레거시 시스템의 검색을 통한 결과를 수집하여 XMDR에 적용시키고, 사용자에게 제공하는 수집기능을 수행한다.

㉢ 검색 에이전트 (Retrieve Agent)

변환 수집 에이전트에 의해 생성된 XML_Query를 이용해 레거시 시스템의 데이터베이스의 XMDR 래퍼에 표준 질의를 전송하는 역할을 수행한다. 질의 전송 후 수행된 결과는 변환 수집 에이전트에 의해 수집되어 임시저장소인 Temp Table에 저장된다. 질의 결과가 없을 경우 관련 키워드를 이용하여 다음 검색이 이루어 질 수 있도록 한다. 인터페이스에 의해 입력된 사용자의 키워드는 알고리즘 3.3의 검색 알고리즘에 의해 생성되어 각 레거시 시스템으로 전송된다. 전송된 질의는 레거시 시스템에서 설치된 XMDR 래퍼에 의해 매핑되어 질의를 수행한다.

```

Void retrieve_Agent() {
...
    int i
    for (i = 0; i =Last XMDR_stack; i++)
    {
        // stack에 저장된 레거시 시스템의 Location
        LOCID = XMDR_stack_legacy_id
        // XMDR을 검색하여 질의문 생성
        convertquery(LOCID, input keyword, return_sql)
        // convertquery()를 이용하여 반환된 질의문
        XML_query = return_sql;
    }
...
//질의문 전송
sendresult(XML_query, LOCID, URL);
}
}
    
```

(알고리즘 3.3) (검색 에이전트 알고리즘)
(Algorithm 3.3) (Retrieval Agent Algorithm)

㉣ 조정 에이전트 (Coordinator Agent)

조정 에이전트(Coordinator Agent)는 파서를 통해 파싱된 키워드, 상품명, 테마의 결과값을 저장하고, 키워드 값을 Knowledge Base에 구축된 전문가 개념망과 일치하는 키워드가 있는지 확인하고, 일치하는 키워드가 있다면 카테고리를 생성한다. 알고리즘 3.4는 카테고리를 생성하기 위해 전문가 개념망과 키워드 개념망의 키워드들을 비교하여 처리하는 과정에 대한 알고리즘이다. 그리고 조정 에이전트는 상품명과 키워드 개념망 사이의 속성을 이용하여 관계를 확인하는 역할을 한다. 그림 3.11은 상품명과 키워드 개념망 사이의 관계를 보여주는 XML문서의 일부 내용이다.

```

Analyzer::KeywordConceptNet_Analyze( {
// 파싱된 결과에서 상품을 하나의 일러온다.
while( CArticle OneProduct = GetProduct( m_ParsedResult ) )
{
// 가까운 상품에서 키워드 리스트 생성한다.
CKeywordList KeywordList = OneProduct.GetKeyWordList( );

while( CKeyword Keyword = KeywordList.GetOneKeyword( ) )
{
// 키워드를 점의 횟수를 저장한다.
m_CategoryDictionary.Add( Keyword );
// 전문가 카테고리에 등록된 키워드만 유저 카테고리를 일치 시킨다.
if( m_ExpertCategory.CheckKeyword( Keyword ) )
{
m_ExpertCategory.MatchUserCategory( Keyword );
}
}
}
}
    
```

(알고리즘 3.4) 전문가 개념 망과 키워드 개념 망 비교 알고리즘

(Algorithm 3.4) The Expert Concept Lattice And Keyword Concept lattice of Comparisons

```

->
<- http://www.semanticweb.org/ontologies/2011/11/Ontology1323069217774.owl#10th_Product ->
<- owl:NamedIndividual rdf:about="http://www.semanticweb.org/ontologies/2011/11/Ontology1323069217774.owl#10th_Product"
<- rdf:type rdf:resource="http://www.semanticweb.org/ontologies/2011/11/Ontology1323069217774.owl#Tokyo" />
<- owl:NamedIndividual
<- http://www.semanticweb.org/ontologies/2011/11/Ontology1323069217774.owl#1st_Product ->
<- owl:NamedIndividual rdf:about="http://www.semanticweb.org/ontologies/2011/11/Ontology1323069217774.owl#1st_Product"
<- rdf:type rdf:resource="http://www.semanticweb.org/ontologies/2011/11/Ontology1323069217774.owl#Osaka" />
<- <CityOf rdf:resource="http://www.semanticweb.org/ontologies/2011/11/Ontology1323069217774.owl#1st_Product" />
<- <asKeyword rdf:resource="http://www.semanticweb.org/ontologies/2011/11/Ontology1323069217774.owl#1st_Product" />
<- <asTheme rdf:resource="http://www.semanticweb.org/ontologies/2011/11/Ontology1323069217774.owl#1st_Product" />
<- owl:NamedIndividual
<- http://www.semanticweb.org/ontologies/2011/11/Ontology1323069217774.owl#2nd_Product ->
<- owl:NamedIndividual rdf:about="http://www.semanticweb.org/ontologies/2011/11/Ontology1323069217774.owl#2nd_Product"
<- rdf:type rdf:resource="http://www.semanticweb.org/ontologies/2011/11/Ontology1323069217774.owl#Osaka" />
<- owl:NamedIndividual
<- http://www.semanticweb.org/ontologies/2011/11/Ontology1323069217774.owl#4th_Product ->
<- owl:NamedIndividual rdf:about="http://www.semanticweb.org/ontologies/2011/11/Ontology1323069217774.owl#4th_Product"
<- rdf:type rdf:resource="http://www.semanticweb.org/ontologies/2011/11/Ontology1323069217774.owl#Tokyo" />
<- owl:NamedIndividual
<- <-
    
```

(그림 3.11) (상품과 키워드 개념망의 연결을 나타내는 XML문서)

(Figure 3.11) (Product and Keyword Concept lattice of XML document Representing the Connection)

3.2.3 레거시 시스템(Legacy System)

레거시 시스템은 제안된 본 시스템에서 실 데이터 계층으로 규정하고 있는 레거시 시스템 자체적으로 유지 관리되는 데이터베이스와 웹서버에서 제공되는 표준과 레거시 시스템의 데이터베이스 스키마와의 매핑 내용을 유지하여 표준을 레거시에 적합하게 변경하고, 레거시의 내용을 표준의 형태로 변환하는 역할을 수행하는 XMDR 래퍼로 구성된다. 여기서는 제안된 시스템에 의해 제공 되는 XMDR 래퍼에 대해서 다루고자 한다.

- XMDR 래퍼 (XMDR Wrapper)

XMDR 래퍼는 레거시 시스템의 데이터베이스 정보를 웹서버에서 제공하는 XMDR 표준과 매핑정보를 유지하고, 표준 질의를 해당 레거시 시스템의 데이터베이스에 적합한 질의로 변환하는 역할을 담당하는 계층이다. XMDR 래퍼는 레거시 시스템 자체의 변화는 없이 웹 서버에 레거시 시스템의 데이터베이스 정보를 등록하고, XMDR의 MSO와 ML을 갱신시킨다. 갱신된 XMDR에서 매핑정보를 가진 XMDR 래퍼를 레거시 시스템에 설치한다. 매핑 정보를 XMDR 래퍼에서 관리함으로써 웹서버에서 전송되는 표준 질의를 레거시 시스템에 적합하게 변환하여 질의를 수행하고, 수행된 결과는 다시 표준으로 변환하게 한다. 이러한 수행 작업을 함으로써 사용자에게는 요청한 정보를 제공하고, 레거시 시스템의 정보를 유출하지 않고 정보를 제공하게 된다. XMDR 래퍼는 알고리즘 3.5에 의해 수행된다.

```

procedure convertquery(char theme_id, char input_keyword, char return_sql)
{
...
//XMDR 래퍼에 내포된 메타정보 검색
XMDR_LOC_stack = findeXMDR(theme_id)
return sql = 'SELECT' + mapping('MSOID, XMDR_LOC_stack_legacy') +
'AS' + 'FIELD1'+ ...
'FROM' + XMDR_LOC_stack LOCHAME + XMDR_LOC_stack TABLENAME +
'WHERE' +mapping('MSOAME', XMDR_LOC_stack_legacy_id)
// XMDR의 표준 MSOID에 해당하는 메타정보를 이용한 질의문 생성
...
}
procedure mapping(char MSOID, char legacy_id, char MSO_attribute)
{
...
// MSO에 대응되는 메타의 속성명 검색
XMDR1_STAND_stack = findeXMDR(legacy_id, MSOID)
MSO_attribute = XMDR1_STAND_stack ontstand
...
}
    
```

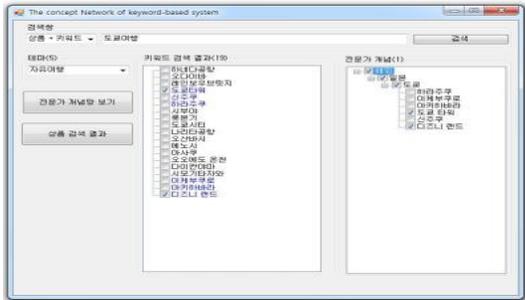
(알고리즘 3.5) XMDR 래퍼의 질의 변환

(Algorithm 3.5) XMDR Wrapper Query Translation

- 실 데이터(Real Data)

실제 데이터 계층은 레거시 시스템들이 관리하고 검색의 대상이 되는 실제 데이터베이스들에 의한 계층으로, 이 계층은 XMDR 래퍼를 통해 검색되며, 웹서버의 영향 없이 독립적으로 각 레거시 시스템들에 의해 관리된다.

그림 3.2의 개념망의 관계 표현과 같이 키워드 개념 망 'B'와 상품, 그리고 테마 개념 망'C'와 상품에 연결되어 있는 관계 속성을 이용하여 여행 상품명과 키워드, 테마 분류, 상품의 출발일과 항공사가 검색결과로 출력된다.



(그림 4.3) 키워드 및 테마 선택화면

(Figure 4.3) Select Display of Keyword and Theme

ProductID	ProductTitle	Keyword	Theme	date	Airline
1232589	도쿄 4일 - 자유	도쿄대행 도쿄...	자유여행	2011.12	JAL
1234003	도쿄 3일 - 자유	도쿄대행 도쿄...	자유여행	2011.12	JAL
3368711	도쿄 5일 - 자유	도쿄대행 도쿄...	자유여행	2011.12	JAL
6469978	도쿄특가상품 3.	특가 도쿄 디즈...	자유여행	2011.12	JAL
6657893	(일본-미국)디즈...	도쿄대행 5일...	자유여행	2011.12	JAL
9782138	도쿄/하코네 4/...	도쿄 디즈니행...	자유여행	2011.12	JAL
1232655	도쿄/디즈니하...	도쿄대행 디즈...	자유여행	2011.12	대한항공
1516519	소통가 도쿄 6.	도쿄 소통가 디...	자유여행	2012.12	대한항공
5369872	도쿄/디즈니하...	도쿄대행 디즈...	자유여행	2012.1	대한항공
1208333	(미국)다문화가...	도쿄대행 자유...	자유여행	2011.12	아시아나항공
1215786	도쿄/시즈오카...	도쿄 시즈오카...	자유여행	2011.12	아시아나항공
3426547	도쿄/가부리자...	도쿄 디즈니하...	자유여행	2011.12	아시아나항공

(그림 4.4) 최종 선택 결과화면

(Figure 4.4) Final Result Display

4.2 비교 분석

제안시스템은 키워드를 검색해 줄 뿐 아니라 전문가 개념 망과 키워드 개념 망을 이용하여 검색한 키워드와 관련된 키워드 리스트들을 보여주며, 그 키워드의 카테고리화 상의 테마를 사용자에게 제공한다. 제안 시스

(표 4.1) 제안된 시스템과 기존 시스템의 비교

(Table 4.1) Comparison of proposed and existing systems

	계층적 트리구조 검색시스템[22]	시맨틱웹 기반 검색시스템[23]	제안 시스템
통합검색	지원 안 됨	지원	지원
	하나의 부모만을 갖는 트리구조라 통합검색 불가	분산 정보 자원을 의미적으로 통합	개념망을 이용하여 분산 정보자원을 통합
시맨틱 검색	지원 안 됨	지원	부분 지원
	시맨틱 검색을 지원 안함	시맨틱 검색 지원	개념망을 이용한 키워드 기반의 검색이지만 개념망간의 관계를 통해 의미적으로도 사용이 가능
사용용이성	높음	보통	높음
	키워드만 입력하여 검색하기 때문에 편리함	시맨틱웹의 UI를 이해해야 편리함	사용자는 키워드만 입력하면 다른 개념망을 추출할 수 있어 편리
연관검색지원	지원 안 됨	부분 지원	지원
	시맨틱검색이 지원되지 않기 때문	시맨틱 검색을 통해 부분 지원	여러 종류의 개념망을 이용한 연관검색가능
검색 효율성	보통	보통	효율적
	특정 검색에 대해서는 효율적으로 찾을 수 있다	키워드 기반의 연관 검색이 가능	개념망을 이용한 키워드 검색이 가능하므로 효율적
검색 적합성	적합지 않음	보통	적합함
	시맨틱 검색이 지원되지 않아 적합성이 떨어짐	아직 완벽한 시맨틱 검색이 아니므로 적합성이 보통	개념망들이 서로 협업한 후 필터링을 거치므로 사용자의 의도에 맞게 적합한 검색이 가능
수정 용이성	용이하지 않음	보통	용이함
	잘못된 검색어를 입력했을 때 수정할 수 있는 유연성이 떨어짐	연관검색어를 통해 검색어 수정이 가능	추출한 다른 개념망을 이용하여 검색어 및 테마 수정가능

템과 같이 기존의 검색 시스템들의 기능을 표 4.1에서와 같이 7가지 항목(통합검색, 시맨틱 검색, 사용용이성, 연관검색지원, 검색 효율성, 검색적합성, 수정용이성)을 통해 비교 분석하였다. 표 4.1은 제안된 시스템과 기존시스템과의 중요기능을 비교한 것이며 기능의 일부를 지원하나 만족하는 경우에는 ‘부분 지원’이라 하였다.

5. 결 론

레거시 시스템마다 카테고리의 분류 기준이 다르고 검색 방식의 차이가 있기 때문에 사용자는 통합된 검색의 결과를 얻기가 어려웠다. 하지만 본 논문에서 제안한 시스템은 레거시 시스템들 간의 이질성을 해결하기 위해 XMDR을 사용하여 표준을 이용한 통합검색이 가능하도록 하였으며, 다수의 개념 망을 사용하여 검색의 효율성을 높이도록 하였다. 제안 시스템의 적용사례에서 살펴보면 키워드 중심의 개념 망과 전문가 개념 망, 테마 망을 이용하여 협업적 검색을 수행하였다. 제안된 시스템은 검색 키워드간의 관련성을 통계적으로 계산한 키워드 개념 망을 구성하여 단순 키워드 검색뿐만 아니라 개념 망을 통한 전문가 개념 및 통계적 수치를 확인할 수 있도록 지원한다. 이는 계층적인 트리 구조의 한계를 넘어 사용자 질의가 가지는 의미로부터 키워드간의 관계에 의한 검색을 가능하게 함으로써 단일 키워드뿐만 아니라 키워드간의 관계에 의해 통합된 키워드 검색을 지원한다. 또한 개념 망을 이용하여 사용자에게 검색 결과를 제공함으로써 기반지식이 없는 사용자로 하여금 관련 키워드를 통해 사용자의 검색목적을 달성할 수 있게 도와준다. 또한 사용자의 요청과 일치하는 검색의 결과가 출력되지 않을 경우, 키워드 개념망을 이용한 관련 키워드를 제공하여 검색을 할 수 있도록 한다. 그리고 분야 개념 망을 이용하여 결과를 나타냄으로써 검색 키워드와 연결된 상품의 테마를 볼 수 있으며, 사용자가 원하는 테마의 상품을 쉽게 찾을 수 있다. 향후 연구 과제로는 여행정보 시스템뿐만 아니라 개념 망 기반의 통합검색 시스템으로 보완하고 확장할 필요가 있다. 또한 여행 테마의 정보와 전문가 카테고리의 관계에 가중치를 부여하여 키워드 연관성과 신뢰성을 높이는 연구가 요구된다. 즉 서로 다른 개념 망을 이용하여 키워드 개념망의 키워드 연관성에 대한 신뢰성을 높이는 연구가 진행되어야 할 것이다.

참고문헌(Reference)

- [1] Vo Ngoc Anh, Owen de Kretser, Alistair Moffat, "Vector-Space Ranking with Effective Early Termination" In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.35-42, 2001, <http://dx.doi.org/10.1145/383952.383957>
- [2] Ionescu, D, Kossmann, D, and Manolescu, L, "Integrating Keyword Search into XML Query Processing," Computer Networks, Vol.33, No.1-6, pp.119-135, 2000. [http://dx.doi.org/10.1016/S1389-1286\(00\)00069-4](http://dx.doi.org/10.1016/S1389-1286(00)00069-4)
- [3] Ji-Hui Im, Young-Jun Bae, Ho-Seop Choi, Cheol-Young Ock, "A Measure of Semantic Similarity and its Application in User-Word Intelligent Network" KCC, Vol.34, No.01, pp.0189-0193, 2007.06
- [4] Peter D. Turney, Patrick Pantel, "From Frequency to Meaning : Vector Space Models of Semantics", Journal of Artificial Intelligence Research 37 (2010) 141-188, <http://dx.doi.org/10.1613/jair.2934>
- [5] Ji-Rong Wen, Ni Lao, Wei-Ying Ma, "Probabilistic Model for Contextual Retrieval" In IJCAI 05 : Proceedings of the 19th international joint conference on Artificial intelligence(2005), pp.1034-1041, <http://dx.doi.org/10.1145/1008992.1009005>
- [6] Wille, R, "Restructuring lattice theory: an approach based on hierarchies of concepts", In:Ivan Rival(ed.), Ordered sets, Reidel, Dordrecht Boston, pp.445-470, 1982, http://dx.doi.org/10.1007/978-94-009-7798-3_15
- [7] Ganter, B, Wille R, "Formal Concept Analysis: Mathematical Foundations", Heidelberg, Springer, 1999. <http://dx.doi.org/10.1007/978-3-642-59830-2>
- [8] M. Bain, "Inductive construction of ontologies from Formal Concept Analysis", Australian Conference on Artificial Intelligence, 2003, pp. 88 - 99. http://dx.doi.org/10.1007/978-3-540-24581-0_8
- [9] R. Belohlávek, J. Dvůrák, J. Outrata, "Fast factorization of Concept Lattices by similarity: solution and an open problem", in: V. Snásel, R. Belohlávek (Eds.), Proceedings of Concept Lattices and their Applications (CLA), Ostrava, Czech Republic, 2004, pp. 47 - 57
- [10] W.W. Cohen, "Data integration using similarity joins and a word-based information representation language", ACM

- Transactions on Information Systems, 18 3 (2000), pp. 288 - 321.
<http://dx.doi.org/10.1145/352595.352598>
- [11] S. Hwang, H.G. Kim, H.S. Yang, A FCA-based ontology construction for the design of class hierarchy, International Conference on Computational Science and its Applications (ICCSA) (3), 2005, pp. 827 - 835, http://dx.doi.org/10.1007/11424857_90
- [12] Kim, Mi-Hye, "Adaptive Learning System based on the Concept Lattice of Formal Concept Analysis", Journal of the Korea Contents Association, Vol. 10 no. 10 (201010), p. 479-493.
- [13] XMDR, <http://www.xmdr.org/>
- [14] ISO/IEC-IS 11179, "Information technology Specification and standardization of data elements", 2003.
- [15] Keck, K.D., McCarthy, J.L, "XMDR: Proposed Prototype Architecture Version 1.01", February, 2005.
- [16] Jin Ma, Seok-Jae Moon, Gye-Dong Jung, Young- Keun Choi, "Design and Implementation of XMDR based on OGSA-DAI System for Data Integration retrieval", Proceedings of KIPS Conference 2009 VOL. 16 NO. 02, pp 0173 ~ 0174 (2009.11)
- [17] Jin Ma, Seok-Jae Moon, Gye-Dong Jung, Young- Keun Choi, "Design of Multi-agent system based on the P2P Networks using Query Rewriting", The 34th KIPS Fall Conference 2010. VOL. 17 NO. 02pp 1780 ~ 1783 (2010.11)
- [18] T.Berners-Lee, The Semantic Web, Scientific American, 2001.
- [19] Antoniou and van Harmelen. A Semantic web Primer, The MIT Press, Cambridge, 2004.
- [20] F.Manola and E.Miller, RDF Primer (<http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>), W3C (MIT, ERCIM, Keio), 2004.
- [21] M.K.Smith, C.Welty, and D.L.McGuinness, OWL Web Ontology Language Guide (<http://www.w3.org/TR/owl-guide/>), W3C (MIT, ERCIM, Keio), 2004.
- [22] Sung-Kwan Kang, Jung-Hyun Lee, "Light-Ontology Classification for Efficient Object Detection using a Hierarchical Tree Structure", Journal of Digital Convergence, Vol.10, No.10, 2012.11 pp. 215 - 220
- [23] Yu Guo, Jie Liu, "Model Design of Semantic Website Construction", JOURNAL OF SOFTWARE, VOL. 9, NO. 8, AUGUST 2014, pp.2142 - 2147, <http://dx.doi.org/10.4304/jsw.9.8.2142-2147>

● 저 자 소 개 ●



마 진 (Jin Ma)

2010년 광운대학교 컴퓨터소프트웨어학과(공학사)
2012년 광운대학교 대학원 컴퓨터과학과(공학석사)
2012~2015.04 비스텔(BISTel) 선임연구원
2015.05~현재 한국과학기술정보연구원 연구원
관심분야 : 분산시스템, 빅 데이터
E-mail : majin0606@gmail.com



전 인 호 (In-Ho Jeon)

2009년 광운대학교 전자공학과(공학사)
2009~현재 광운대학교 대학원 전자공학과(박사과정)
2013~현재 한국과학기술정보연구원 연구원
관심분야 : e-Science, 무선통신
E-mail : inojeon@kisti.re.kr



최 영 근 (Young-Keun Choi)

1980년 서울대학교 수학교육학과(이학사)
1982년 서울대학교 대학원 계산통계학과(이학석사)
1989년 서울대학교 대학원 계산통계학과(이학박사)
1983~현재 광운대학교 컴퓨터소프트웨어학과 교수
관심분야 : 객체지향설계, 분산시스템
E-mail : ygchoi@kw.ac.kr