

# FCA 개념 망 기반 개인정보관리

## Personal Information Management Based on the Concept Lattice of Formal Concept Analysis

김 미 혜\*  
Kim, Mihye

### 요 약

개인정보관리 시스템의 궁극적인 목표는 개인이 필요로 하는 정보를 수집하고 처리하여 보다 간편하고 효율적인 방법으로 정보를 검색할 수 있도록 조직화하여 관리하는데 있다. 그러나 기존의 개인정보관리 시스템은 자료 저장을 위해 전통적인 계층적 디렉터리 모델을 사용하고 있기 때문에 효율적인 정보관리 및 검색에 한계가 있으며 정보의 연관 관계에 의한 검색이 어렵다는 문제점을 가진다. 이러한 문제점들을 개선하기 위해 본 논문에서는 개인이 필요로 하는 문서를 웹상에서 쉽게 관리하고 유지할 수 있는 더불어 관리된 문서를 쉽게 검색할 수 있는 개인정보관리 메커니즘을 기계학습의 한 기법이 Formal Concept Analysis의 개념망 모델을 이용하여 제안한다. 실험 결과는 제안된 방법이 계층적인 트리 구조의 단점을 극복하고 제한된 검색을 넘어 개념들 사이의 연관 관계에 의한 검색을 지원할 뿐만 아니라 검색 성능에 있어 계층적인 구조를 기반으로 하는 시스템보다 더 유용한 방법임을 보였다.

### Abstract

The ultimate objective of Personal Information Management (PIM) is to collect, handle and manage wanted information in a systematic way that enables individuals to search the information more easily and effectively. However, existing personal information management systems are usually based on a traditional hierarchical directory model for storing information, limiting effective organization and retrieval of information as well as providing less support in search by associative interrelationship between objects (documents) and their attributes. To improve these problems, in this paper we propose a personal information management model based on the concept lattice of Formal Concept Analysis (FCA) to easily build and maintain individuals' own information on the Web. The proposed system can overcome the limitations of the traditional hierarchy approach as well as supporting search of other useful information by the inter-relationships between objects and their attributes in the concept lattice of FCA beyond a narrow search.

☞ Keyword : Personal Information Management, Concept lattice, Browsing

## 1. 서 론

현대사회를 일컬어 정보시대라 한다. 오늘날의 조직이나 기업은 물론 개개인도 인터넷의 발달과 함께 끊임없이 생성되는 수많은 정보 속에 놓여져 있다. 그러나 방대한 비정형 정보의 홍수 속에 진정 원하는 정보도 찾기 어렵지만 얻은 정보를 효율적으로 관리하고 필요시에 검색하여 재사

용하기 또한 그리 쉬운 일은 아니다. 이러한 정보화 시대에 부응하여 개인정보관리(PIM: Personal Information Management)에 대한 연구가 데이터 관리 커뮤니티의 중요 관심 분야로 부각되고 있다. 몇몇 개인정보관리 시스템에 대한 연구가 진행되고 있지만 대부분의 정보 관리 응용시스템은 도메인 정보를 거의 사용하지 않는 일반적인 검색엔진의 접근방법과 수작업으로 만들어지는 계층적인 분류체계의 디렉터리나 혹은 파일 시스템의 접근방법을 주로 많이 이용하고 있다. 팔목할만 한 검색 엔진의 발전에도 불구하고 이러한 방

\* 정 회 원 : 대구가톨릭대학교 컴퓨터교육과 교수  
mihyekim@cu.ac.kr(제1저자)  
[2005/08/03 투고 - 2005/09/16 1차 심사 - 2005/11/09  
2차 - 2005/11/15 심사완료]

법들은 대부분 통계적인 모델을 기반으로 한 구문해석에만 의존하고 있어 구조화된 지식관리 및 효율적인 검색 성능 향상에 한계가 있음이 드러나고 있다.

한 대안적인 방법으로 전형적인 검색엔진의 문제점들을 해결하고 보다 효율적인 지식관리를 위한 새로운 접근 방법에 대한 연구가 활발히 진행되고 있다. 이 방법은 온토로지를 기반으로 하여, 웹 내용에 기계가 처리 가능한 의미정보를 내장함으로써 웹 자원을 강화하여 보다 정확한 정보를 허용하고자 하는데 목적이 있다. 이러한 방법은 보다 체계적으로 지식을 관리할 수 있게 하고 의미론적으로 마크업된 온토로지들에 근거하여 보다 정확한 정보를 제공할 수 있을 뿐만 아니라 온토로지들의 상호 관련성에 근거하여 내제되어 있는 새로운 지식을 자동으로 추론할 수 있어 보다 향상된 정보 제공이 가능하다는 장점이 있다. 또한 웹상에서 지식의 공유와 재사용, 상호 운용성에 있어 온토로지의 기여는 대단할 것으로 본다. 하지만 이 방법은 온토로지 생성 및 구성, 지식의 불일치성, 지식 획득의 병목 현상, 사용자 질의어 인터페이스 등 해결해야 할 여러 이슈들이 아직 많이 남아 있다[1]. 더불어 온토로지에서 기인하는 장점을 충분히 활용하기 위해서는 사용자가 온토로지의 개념을 상당 부분까지 이해하고 있어야만 한다는 전제 조건이 따른다. 하지만 온토로지 개념은 최종 사용자가 이해하기에는 그리 쉬운 개념은 아니므로 따라서 사용자의 편리성 및 가용성(usability) 측면에서 주요한 논쟁의 여지를 내포하고 있으며 아직도 학계에서 “온토로지란 무엇인가” 라는 주제가 쟁점화 되고 있다는 사실이 이를 뒷받침하고 있다.

본 논문의 목적은 궁극적으로는 온토로지 접근 방법과 목적은 같아지지만, 보다 유연하고 자유로운 문서 주석(annotation)에 기초하여 사용자 중심으로 접근한 보다 개방적이고 쉬운 개인 정보관리 및 검색 시스템에 목적을 두고 있다. 즉, 미리 정의된 분류체계 시스템을 기반으로 시작하

여 특정분야의 정보나 혹은 개인의 정보를 관리하기보다는 사용자가 문서의 의미를 자유롭게 정의하고 관리해 나감에 따라 적용분야에 합당한 시스템이 점진적으로 생성 발전되어 갈 수 있는 방법에 목적을 두고 있다. 이러한 견해는 어떠한 지식 구조(knowledge structure)든지 시스템이 전개되어 감에 따라 발전 진보되어 가는 구성체(construct)라는 견해[2]에 기인한다. 다시 말해 지식은, 언제나 형성 중에 있으며 불완전하고 관련된 시스템과 지속적이고 반복적인 대화를 통해 발전되어 가는 속성을 지닌다는 것이다[3].

본 논문에서는 개인이 필요로 하는 문서를 웹상에서 쉽게 관리하고 유지할 수 있는 더불어 관리된 문서를 간편하고 쉽게 검색할 수 있는 새로운 접근방법의 개인정보관리 메커니즘을 기계학습의 한 기법인 FCA(Formal Concept Analysis) [4,5] 모델을 이용하여 제안한다. 제안된 방법은 사용자가 주어진 인터페이스를 통해 문서의 의미를 자유롭게 정의하고 관리해 나감에 따라, FCA의 개념망에 상응하여 적용 분야에 합당한 망(lattice) 구조로 이루어진 분류체계(taxonomy) 시스템이 자동적이고 점진적으로 생성 발전되어 갈 수 있도록 한다. 더불어 이 개념망을 정보검색을 위한 브라우징 메커니즘으로 활용함으로써 계층적인 트리 구조의 단점을 극복하고 제한된 검색을 넘어 사용자가 필요한 문서를 검색할 수 있도록 지원할 뿐만 아니라 분류체계 내에 존재하는 개념들 사이의 연관관계에 의한 검색 또한 가능하도록 한다.

본 논문의 구성은 다음과 같다. 제2장에서는 관련 연구로서 웹 디렉터리 시스템을 포함한 상용화된 웹 검색 엔진의 기법들과 온토로지 접근 방법을 소개하고 장단점을 논의한다. 제3장에서는 제안된 시스템의 주요 방법론인 FCA 모델에 대해 소개한다. 제4장에서는 FCA 모델을 개인정보관리 시스템에 적용시키는 방안과 구현한 시스템에 대해서 설명한다. 제5장에서는 실험결과를 바탕으로 제안된 시스템의 장점 및 성능평가 결과를 보인다. 제6장에서는 본 논문을 요약하고 결

론과 함께 향후 연구를 제시한다.

## 2. 관련연구

문서관리의 궁극적인 목적은 사용자가 좀 더 쉽게 원하는 정보를 찾을 수 있도록 보다 효율적인 방법으로 문서들을 조직화하여 관리하는데 있다. 상용화된 웹 검색 엔진은 크게 두 가지 방법으로 문서를 색인 화하여 관리한다. 첫 번째 방법은 주로 로봇이나 웹 탐색기를 이용해 웹상의 정보를 정기적으로 수집하고(크로링: crawling) 수집된 문서들은 문서에 포함되어 있는 정규화된 단어들을 바탕으로 벡터공간 모델이나 확률 모델 등을 이용하여 색인화 된다. 새로운 웹 기술을 기반으로 하나의 웹 사이트가 다른 웹 페이지로부터 링크되어 있는 빈도수에 따라 순위를 부여하는 하이퍼링크 정보를 이용하는 웹 검색 엔진도 있다 (Google: <http://www.google.com>과 Teoma: <http://www.teoma.com>). 색인화된 문서들은 사용자 질의어에 대응하여 시스템과의 상호작용을 통해서 검색된다. 이 방법은 일반 사용자 누구나 쉽게 접근할 수 있는 간편한 질의어 인터페이스를 제공하기 때문에 사용자에게 원하는 정보만 항상 제공해 줄 수 있다면 가장 이상적인 방법이라 여겨진다. 팔목할 만 한 검색 엔진의 발전에도 불구하고 많은 경우 특정 질의어에 대한 검색 결과는 매우 실망스럽고 검색을 위한 적절한 질의어를 재 정의하는 것 또한 그리 쉬운 일이 아니다.

두 번째 방법은 이미 정의된 분류체계 시스템이나 웹 디렉터리 시스템에 문서의 특성을 표현하는 범주를 결정하여 문서를 분류체계의 계층적인 구조 안에 조직화 시키는 것이다. 이는 선-클러스터링(pre-clustering) 기법으로 Yahoo나 Google의 디렉터리 시스템이 대표적인 경우이다. 사용자는 조직화된 구조를 브라우징 하여 원하는 정보를 찾을 수 있다. 또 하나의 웹 클러스터링 방법은 후-클러스터링(post-clustering) 기법으로[6] 일반적인 검색 엔진으로부터 사용자 질의어에 대응하여 검색된

정보 항목들로부터 각 범주별 특성을 자동적으로 추출하여 분류체계 시스템을 동적으로 형성하고 사용자는 이 분류시스템을 브라우징하여 원하는 정보를 찾을 수 있다. Vivisimo(<http://vivisimo.com>)나 WiseNut(<http://www.wisenut.com>) 과 같은 검색 엔진이 이러한 기법을 사용하고 있는 대표적인 경우이다. 이 방법은 검색 결과를 분류체제로 정리하여 사용자에게 제시함으로써 일반적인 웹 검색 시스템의 단점인 선형적인 리스트 안에 제시된 방대한 양의 검색 결과로부터의 과부하를 줄이고 좀 더 신속하고 쉽게 원하는 정보에 접근할 수 있도록 한다[6]. 클러스터링을 이용하는 브라우징 시스템의 장점은 사용자가 신속하고 쉽게 도메인 지식을 획득하고 탐색할 수 있도록 지원하는데 있다[7]. 따라서 이러한 장점에 기인하여 대부분의 개인정보관리 시스템 또한, 클러스터링 페러다임을 따른 브라우징 시스템을 이용하여 관리되고 검색된다.

그러나 대부분의 전통적인 개인정보 관리 응용 시스템은 수작업을 통한 선-클러스터링 기법에 근거하여 필요에 따라 수작업으로 새로운 클러스터(cluster) 경로를 추가해가며 전형적인 트리 구조인 디렉터리 색인에만 의존하고 있어 다른 정보와의 연관관계에 의한 검색이 어렵다. 더불어 만약에 사용자가 잘못된 경로를 선택하여 검색을 진행한 경우에는 다른 클러스터로 네비게이션할 수 있는 메커니즘이 없으므로 단순히 처음부터 다시 시작해야만 한다. 또 하나의 중요한 문제점은 미리 정의된 분류체계에서는 문서를 대표할 수 있는 가장 적절한 클러스터 항목을 결정한 후 그 계층적인 클러스터 안에 문서를 색인하기 때문에 카테고리 불일치(category mismatch) [8] 현상이 발생할 수 있다. 수작업으로 새로운 클러스터를 생성해가며 관리하는 경우 또한 클러스터 간의 불일치 현상은 언제든지 발생할 수 있다.

한 대안적인 접근방법으로 검색 성능 향상을 목적으로 하는 보다 효율적인 지식관리 방법론이 각광을 받고 있다. 이러한 연구는 현재 웹 자원의 대표적인 표현 언어인 HTML 언어로는 문서

의 정형화된 시맨틱 정보를 기술하는데 어려움이 있으며, 이러한 한계를 극복하기 위해서는 새로운 웹 표현 언어를 사용해 기계가 처리 가능한 의미 정보(machine processable semantics)를 추가하여 웹 자원을 강화하여야 한다는 것이다. HTML언어는 단지 레이아웃이나 포매팅, 하이퍼링크 기술에 초점을 맞춘 웹 문서에 의미 정보를 부여할 수 없으므로 웹 자원에 대한 정형화되고 명시적인 시맨틱 정보들(즉, 온토로지들)을 개발하고 RDF/S, OIL, DAML, DAML+OIL 등과 같은 온토로지 표현 언어들을 이용해 개발된 온토로지를 기반으로 웹에 있는 모든 자료들이 분명한 의미로 주석되어야 한다는 것이다. 일반적으로 이 방법은 지식 공학자나 온토로지 개발자들이 특정분야에 대한 표준화된 온토로지들을 미리 정의의 생성하고 사용자들은 정의된 온토로지를 바탕으로 온토로지 언어를 이용하여 웹 자원을 주석하고 관리한다. 자연언어처리 기술을 적용한 자동마크 업(mark-up)에 대한 연구도[9, 10] 진행 중에 있지만 이것은 장기적인 목표라 여겨진다. 월드와이드웹 표준화 단체인 W3C(World Wide Web Consortium)에 의해 형성된 웹 온토로지 작업 그룹에서는 시맨틱 웹의 표준화된 온토로지 언어로서 OWL(Ontology Web Language)을 제시하였다. 이러한 연구를 주도하는 대표적인 그룹으로서는 시맨틱 웹 커뮤니티 포털(Semantic Web community portal) 과 W3C의 웹 온토로지 작업 그룹(Web Ontology Working Group) 등과 같은 단체가 있으며 이러한 활동의 좀 더 구체적인 사례로는 KA<sup>2</sup> initiative[11]와 Ontoshare[12] 있다.

앞서 언급한 바와 같이 이러한 환경 하에서의 검색 시스템은 웹 데이터에 내장된 의미 정보에 근거하여 보다 정확한 정보 검색 및 의미 정보의 연관관계에 의한 새로운 지식의 자동 추론이 가능하여 보다 복잡하고 한층 높은 탐색 기능을 지원할 수 있지만 온토로지의 구성, 지식의 불일치성, 사용자 질의어 인터페이스 및 가용성 등 해결해야 할 여러 이슈들이 남아 있다[1,13]. 이 방

법은 온토로지의 구조(ontology structure)를 문서 검색을 위한 브라우징 메커니즘으로 제공하지만 [11] 온톨로지 구조 또한 전형적인 디렉터리 시스템과 동일한 계층적인 트리 구조를 이용한다. 물론, 그래프 구조로 표현되는 RDF 메타데이터의 자원들을 하이퍼링크로 표현한 새로운 시맨틱 웹 브라우저 기법에 대한 연구[13]와 시맨틱 검색[14]에 대한 연구가 활발히 진행되고 있다.

### 3. Formal Concept Analysis(FCA)

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. 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 객체집합의 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[4]. FCA은 자료 분석을 위한 여러 응용 분야에 적용되었으며 많은 연구들이 FCA의 개념망을 정보 검색에 적용 발전시켰다[15-18]. FCA는 세 개의 기본 구성요소; 배경도(formal context), 개념(formal concept), 개념망(concept lattice)으로 이루어진다. 여기에서, 형용사 'formal'이라는 단어는 FCA의 모든 기본 요소들이 수학적인 개념으로 다루어지고 있다는 것을 강조하기 위함이다.

#### 3.1 배경도 (Formal context)

배경도는 FCA의 가장 기본적인 자료 구조이며, 세 개의 구성요소  $C=(D, K, I)$ 인 두개의 집

합  $D$ 와  $K$ , 그리고  $D$ 와  $K$ 사이의 관계를 나타내는  $I$ 로 이루어진다.  $D$ 의 원소들은 배경도의 객체,  $K$ 의 원소들은 배경도의 속성이라 부른다.  $dIk$  혹은  $(d, k) \in I$  라 수식은 하나의 객체  $d$ 가 속성  $k$ 와 이항관계  $I$ 에 있음을 나타내기 위해 사용된다. 배경도는 각각의 행에는 객체들, 각각의 열에는 속성들로 이루어진 표 1과 같은 십자도표로 표현된다.

예를 들어, 동물의 세계에서 동물들은 객체에 해당하고 동물들이 가지는 특성들; 날개가 있다, 수영을 할 수 있다, 육식을 한다 등은 속성들에 해당될 수 있겠다. 이 논문에서 제시한 FCA의 구현 분야인 개인정보관리 시스템에서는 관리하고자 하는 문서들은 객체가 되며, 문서의 의미를 나타낼 수 있는 키워드들은 문서의 속성이 된다. 물론 문서의 제목, 저자, 작성한 날짜 등도 문서의 속성이 될 수 있다. 후자의 경우는 FCA의 개념적 축소(conceptual scaling)[5, 19] 기법으로 처리할 수 있으면 본 논문에서는 다루지 않기로 하겠다.

표 1은 개인정보관리 시스템에서의 배경도의 일부분을 보여주고 있다. 이 도표에서 행에 있는 숫자는 객체인 문서들을, 행의 단어들은 속성인 키워드들을 의미한다. 심벌 'X'는 해당 문서가 대응하는 키워드를 가지고 있음을 의미하며 편의상 문서들은 숫자로, 키워드들은 문자로 표시하였다. 표1에 나타난 배경도  $C = (D, K, I)$ 는 문서 집합  $D = \{1, 2, 3, 4, 5, 6, 7\}$ , 키워드집합  $K =$

{Ontology, Semantic Web, Case Study, Ontology Representation Language, OWL, OIL, Web Service, Semantic Web Technology}와 관계집합  $I = \{(1, \text{Ontology}), (1, \text{Semantic Web}), (2, \text{Ontology}), (2, \text{Semantic Web}), (2, \text{Case Study}), \dots, (7, \text{Ontology}), (7, \text{Semantic Web}), (7, \text{Semantic Web Technology})\}$  로 표현할 수 있다.

### 3.2 개념 (Formal concept)

개념은 객체와 속성 사이의 관계를 반영하며 개인정보관리 시스템에서는 문서와 키워드 사이의 관계를 나타낸다. 한 배경도내에 존재 가능한 개념들을 추출하기 위해 다음의 수식들이 정의된다:  $X \subseteq D: X \rightarrow X' := \{k \in K \mid \forall d \in X: (d, k) \in I\}$ ,  $Y \subseteq K: Y \rightarrow Y' := \{d \in D \mid \forall k \in Y: (d, k) \in I\}$ .

위의 수식들을 바탕으로 하나의 개념은  $X \subseteq D, Y \subseteq K, X'=Y$ 와  $X'=Y$ 를 만족하는  $(X, Y)$ 의 쌍으로 정의된다. 예를 들어, 객체집합  $X = \{1, 2, 6, 7\}$ , 속성집합  $Y = \{\text{ontology, semantic web}\}$ 이라 할 때, 객체집합  $X$ 와 이항관계에 있는 속성집합  $X' = \{\text{ontology, semantic web}\}$ 이고 속성집합  $Y$ 와 이항관계에 있는 객체집합  $Y' = \{1, 2, 6, 7\}$  이므로  $(\{1, 2, 6, 7\}, \{\text{ontology, semantic web}\})$  은 표 1에 있는 배경도의 한 개념이 될 수 있다. 객체집합  $X$ 는 개념  $(X, Y)$ 의 extent, 속성집합  $Y$ 는 intent라 부른다.

〈표 1〉 개인정보관리 시스템에 있는 배경도(formal context)의 일부분

	a	b	c	d	e	f	g	h
	Ontology	Semantic Web	Case Study	Ontology Rep. Language	OWL	OIL	Web Service	Semantic Web Technology
1	X	X						
2	X	X	X					
3	X			X	X			
4	X			X		X		
5		X					X	
6	X	X					X	
7	X	X						X

개념망을 형성하기 위해서는 배경도에 존재 가능한 모든 개념들을 찾아내야 한다. 식 3.1과 3.2는 한 배경도에 존재하는 개념들을 추출하기 위해 사용될 수 있는 대표적인 수식들이다. 식 3.1에서는 우선 먼저  $d \in D$ 인 모든 intents  $\{d\}$ '를 결정하고 (식 3.2는  $k \in K$ 인 모든 extents  $\{k\}$ '를 결정), 그 다음 각각의 intent에 대응하는 extent  $X'$ 를 구한다. 표 2는 표 1의 배경도로부터 모든 개념들이 어떻게 추출되는지를 식 3.2에 근거하여 보여주고 있다. 자세한 절차는 다음과 같다.

$$X' = \bigcap_{d \in X} \{d\}' \quad (3.1)$$

$$Y' = \bigcap_{k \in Y} \{k\}' \quad (3.2)$$

**절차1:** 우선 먼저 배경도에 있는 모든 문서들을 포함하는 집합  $\{1,2,3,4,5,6,7\}$ 을 구하여 테이블의 extent항목에 등록 한다 (표 2(a)의 단계1). 그리고 각각의 키워드( $a,b,c,d,e,f,g,h$ )에 대하여 절차 2를 수행한다.

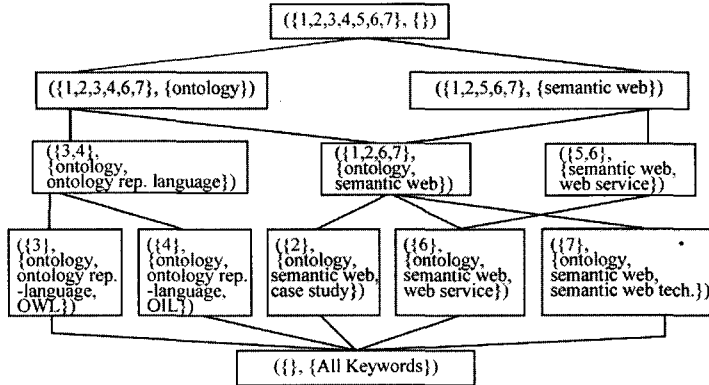
**절차2:** 키워드  $k_i$ , 즉 a를 포함하고 있는 문서집합 extent X를 구한다. 키워드 a를 포함하고 있

는 extent X는  $\{1,2,3,4,6,7\}$ 이다. X가 테이블에 존재하지 않으면 해당 키워드와 함께 하나의 새로운 extent로 테이블에 추가 시킨다 (표 2(a)의 단계2). 그 다음에는 테이블에 등록되어 있는 모든 extent들에 대해 X와의 교집합을 구한다. 구한 교집합이 extent로 등록되어 있지 않으면 그 교집합 또한 해당 키워드의 extent로 테이블에 추가 시킨다. 예를 들면, 표2(a)의 단계2에서 extent X  $\{1,2,3,4,6,7\}$ 을 테이블에 등록시킨 후 테이블에 등록되어 있는 모든 extent들과의 교집합을 구하여야 하고, extent  $\{1,2,3,4,6,7\}$ 이 테이블에 존재하므로 이 집합과의 교집합을 구한다. 구한 교집합은 자신과 동일한  $\{1,2,3,4,6,7\}$ 이므로 테이블에 추가하지 않는다. 교집합이 하나의 새로운 extent로 추가된 경우는 표 2(a)의 단계 3, 5, 8에서 찾아 볼 수 있다. 절차 1과 절차2의 결과가 표 2의 (a)에 나타나 있다.

**절차3:** 그 다음에는 표 2(a)에 있는 각각의 extent X에 대하여 X의 모든 원소와 이항관계에 있는 키워드들의 집합 intent Y ( $Y \leftarrow \{k \in K \mid \forall d \in X: (d, k) \in \mathcal{R}\}$ )를 결정한다. 표 2의 (b)는 절차 3의 결과를 보여주고 있다. 표 1에 있는 배경도에는 총 12개의 개념이 존재함을 알 수 있다.

<표 2> 표 1의 배경도에 존재하는 모든 개념들을 추출하기 위한 절차 (a)  $k \in K$ 인 모든 extents  $\{k\}$  결정. 즉, 각 키워드에 대응하는 extent (문서집합)을 구함 (b) 각각의 extent에 대응하는 intent(키워드집합)을 구함

단계	intent (키워드집합)	extent (문서집합)	단계	intent (키워드집합)	extent (문서집합)
1		{1, 2, 3, 4, 5, 6, 7}	1	{ }	{1, 2, 3, 4, 5, 6, 7}
2	a	{1, 2, 3, 4, 6, 7}	2	{a}	{1, 2, 3, 4, 6, 7}
3	b	{1, 2, 5, 6, 7} {1, 2, 6, 7}	3	{b} {a, b}	{1, 2, 5, 6, 7} {1, 2, 6, 7}
4	c	{2}	4	{a, b, c}	{2}
5	d	{3, 4} { }	5	{a, d} {a, b, c, d, e, f, g}	{3, 4} { }
6	e	{3}	6	{a, d, e}	{3}
7	f	{4}	7	{a, d, f}	{4}
8	g	{5, 6} {6}	8	{b, g} {a, b, g}	{5, 6} {6}
9	h	{7}	9	{a, b, h}	{7}



〈그림 1〉 표 1의 배경도에 대한 개념망(concept lattice)

### 3.2 개념 망(Concept lattice)

개념망은 개념들의 intent들 사이에 원소의 수가 가장 작은 것부터 나열한 FCA의 개념도이다. 개념망을 형성하기 위해서는 모든 개념들 사이의 상·하위(superconcept-subconcept) 개념 관계를 찾아야 한다. 주어진 두 개념  $c_1=(X_1, Y_1)$ 과  $c_2=(X_2, Y_2)$ 가  $(X_1, Y_1) \leq (X_2, Y_2) : \Leftrightarrow X_1 \subseteq X_2 (\Leftrightarrow Y_2 \subseteq Y_1)$ 의 조건을 만족할 때  $c_2$ 는  $c_1$ 의 상위개념,  $c_1$ 은  $c_2$ 의 하위 개념이라 부른다. 예를 들어, 개념  $(X_2, Y_2)=({1,2,6,7}, \{ontology, semantic web\})$ 과  $(X_1, Y_1)=({2}, \{ontology, semantic web, case study\})$ 이 존재할 때  $X_1 \subseteq X_2 (\Leftrightarrow Y_2 \subseteq Y_1)$ 의 조건을 만족하므로 개념  $(X_2, Y_2)$ 은  $(X_1, Y_1)$ 의 상위개념이 되는 것이다. 반대로 개념  $(X_1, Y_1)$ 은 개념  $(X_2, Y_2)$ 의 하위 개념이 된다. 개념망은 모든 개념들 사이의 계층적인 순서, 상·하위 관계를 찾아 라인 다이어그램으로 나타낸 것이다. 그림 1은 표 1의 배경도에 대한 개념망을 보여주고 있다. 각 노드는 하나의 개념에 해당된다. 개념망은 계층적인 트리 구조와는 달리 여러 개의 부모 노드, 즉 여러 경로를 허용한다. 예를 들어, 그림 1에 있는 개념  $(X_1, Y_1)=({1,2,6,7}, \{ontology, semantic web\})$ 은 두개의 상위개념  $(X_2, Y_2)=({1,2,3,4,6,7}, \{ontology\})$ 와  $(X_3, Y_3)=({1,2,5,6,7}, \{semantic web\})$ 을 가진다.

다음 장에서는 개인의 정보가 어떻게 FCA의

개념망으로 관리되고, 새로운 정보가 추가되고 변경됨에 따라 자동적으로 발전 전개되어 가는지를 보인다. 또한 어떻게 개념망이 정보 검색 메커니즘으로 사용되는지를 보인다.

## 4. FCA 개념망 기반 개인정보관리시스템 설계 및 구현

본 논문에서는 개인의 정보를, 예를 들어, 자주 이용하는 웹 사이트, 연구 논문, 학회 정보 등을 부가적인 메타데이터 즉, 제목, 키워드, URL 등을 자유롭게 주석하게 하고 이러한 정보를 바탕으로 기계학습의 한 기법인 FCA의 개념망으로 색인하여 관리하며, 이 개념망을 정보 검색을 위한 브라우징 메커니즘으로 활용함으로써 계층적인 트리 구조의 단점을 극복하고 제한된 검색을 넘어 개념들 사이의 연관 관계에 의한 검색을 지원할 수 있는 새로운 접근 방법을 제안한다.

### 4.1 FCA 개념망 기반 문서 관리

그림 2는 XML로 기술된 연구 논문의 한 예를 보여주고 있다. 사용자는 주어진 인터페이스를 통하여 그림 2에 나타나 있는 메타데이터를 제공할 수 있으며 제공된 정보는 그림 2의 XML 형태로 관리된다. 문서의 제목(title), 키워드(keywords)와

```

<!-- 개인정보: 연구논문 ->
<document>
<title>The Semantic Web: Yet Another Hip?</title>
<keywords>Semantic Web, Ontologies, Semantic Web Technology</keywords>
<url>c:\Tomcat\webapps\search\docs\semanticWeb\DKE41.pdf</url>

<authors>Ying Ding, Dieter Fensel, Michel Klein, Borys Omelayenko</authors>
<publisher>Journal of Data and Knowledge Engineering</publisher>
<year>2002</year>
</document>

```

(그림 2) XML 로 기술된 연구 논문의 한 예제

문서의 위치(URL)는 반드시 주석해 주어야 한다. 키워드들은 문서의 의미를 나타낼 수 있는 중요 단어와 분류 용어 등을 사용하여 자유롭게 지정해 줄 수 있다. 기타 정보(저자, 출판년도 등)는 문서 검색시 검색 결과를 개념적으로 축소(conceptual scaling)하여 볼 수 있는 데이터들이다. 예를 들어, “semantic web”과 관련된 문서 중에서 저자가 “Dieter Fensel”인 문서만을 찾고자 할 때 이용될 수 있다. 앞서 언급한 바와 같이 이 기능에 대해서는 본 논문에서 자세히 다루지 않기로 하겠다.

그림 3은 FCA의 기반 문서 관리의 내부구조를 보여주고 있다. 그림 3의 (a)는 메타데이터에서 지정한 URL에 존재하는 HTML이나 PDF양식의 실제 문서 내용이다. 사용자가 이 문서의 메타데이터(그림 3의 (b) 참조)를 등록하면, 시스템은 메타데이터에 사용된 키워드들과 문서 정보에 근거하여 3장에서 소개한 배경도를 형성하고 그 배경도로부터 존재 가능한 모든 개념들을 추출하여 개념들 사이의 상·하위 관계를 찾아 개념망(그림 3의 (c) 참조)을 구성한다.

그림 3의 (d)는 개념망 안의 색인 구조를 보여주고 있다. 예를 들어, 키워드 “semantic web”과 관련된 문서는 doc1, doc2, doc5, doc6, doc7 이고, 키워드 “ontology & semantic web”과 관련된 문서는 doc1, doc2, doc6, doc7임을 알 수 있다. 따라서 개념망의 하위 레벨로 내려 갈수록 보다 더 특정한 검색 결과를 얻을 수 있음을 알

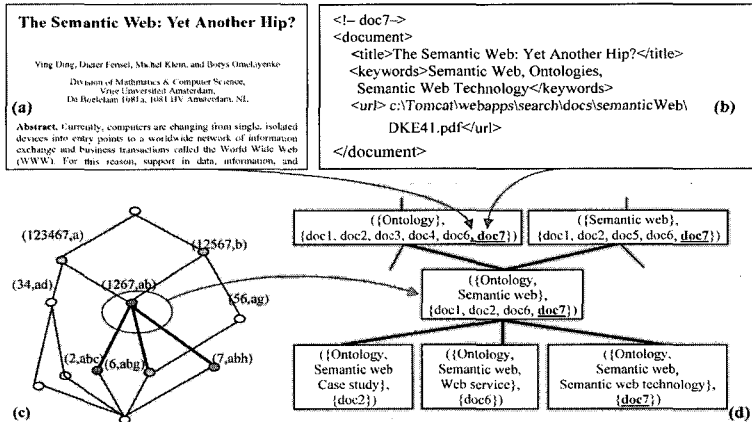
수 있다. 결론적으로, 제안된 시스템에서 이 개념망은 문서를 관리하기 위한 기본 자료구조로 사용될 뿐만 아니라 문서 검색을 위한 브라우징 메커니즘으로도 이용된다. 이 개념망은 새로운 문서가 추가되거나 시스템에 등록되어 있는 문서 정보가 변경될 때마다 점진적이고 자동적으로 재구성 되어진다.

표준화된 클러스터링 스키마의 경우 문서들은 보통, 보다 더 일반화된 개념을 상단에 위치시키는 계층적인 트리 구조 안에 조직화된다. 따라서 루트에서부터 주어진 클러스터까지는 단지 하나의 경로만이 존재하게 된다. 그러나 개념망은 여러 개의 부모 노드 관계를 갖는 lattice 스키마, 즉 여러 경로를 허용하는 그래프 구조이다.

## 4.2 FCA 개념망 기반 문서 검색

제안된 시스템의 주된 검색 메커니즘은 웹을 기반으로 하여 FCA의 개념망을 브라우징 하는데 있다. 사용자는 특정한 질의어를 명시하지 않고 개념망의 최상위 레벨의 노드로부터 시작하여 다른 노드 정점으로 이동하며 개념들의 상호 연관 관계를 탐색하면서 원하는 정보를 검색할 수 있다. 개념망의 구현은 그래프 자체로 도식하여 표현하지 않고 웹 환경에 적합한 하이퍼링크(hyperlink)를 사용하여 가장 가까운 상·하위 개념노드, 즉 직계 부모 노드들과 자식 노드들만을 보여준다. 그림 3의 (c)에 나타나 있는 간단한 구조도를





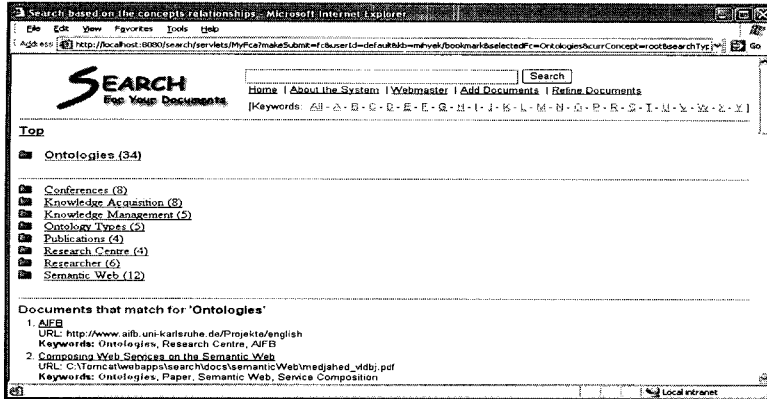
(a) URL에 있는 문서의 내용 (b) 문서의 메타데이터 (c) 개념망(Concept lattice) (d) 개념망의 색인  
 <그림 3> FCA 개념망 기반 문서 관리의 내부구성

이용하여 어떻게 개념망에 대한 네비게이션이 수행되는지를 살펴보고자 하자. 사용자 질의어가 “a”라고 가정할 때 시스템은 질의어와 관련이 있는 문서들 doc1, 2, 3, 4, 6, 7을 검색결과로 보여지게 되고, “a”의 하위 개념인 직계 자식 노드 “ad”와 “ab”를 브라우징 공간에 디스플레이 하게 된다. 만약 이 상태에서 개념 “ab”가 선택되었다면 시스템은 검색결과 공간에 문서 doc1, 2, 6, 7을 디스플레이 하고 이 개념의 상위개념 “a”와 하위개념 “abc”, “abg”와 “abh”를 브라우징 공간에 보여주게 된다. 사용자는 브라우징 공간에 나타나 있는 개념들을 상·하위로 선택하면서 문서 검색을 계속 진행할 수 있고 혹은 개념망의 최상위 레벨부터 다시 시작하여 검색을 진행할 수 있다.

또한 사용자는 전형적인 정보검색의 부울리언(Boolean) 질의 인터페이스를 통해 질의어를 입력하거나 혹은 시스템에서 제공해 주는 키워드 리스트로부터 용어를 선택하여 문서를 검색할 수도 있다. 즉, 시스템은 문서 주석에 사용된 키워드들을 제공하여 준다. 사용자는 코마(,)로 구분된 여러 단어들로 구성된 질의어를 입력할 수 있으며 이는 “AND” 부울리언 연산자로 가정된다. 이러한 경우에 시스템은 일반적인 검색 엔진과 동일한 방

법으로 질의어로부터 먼저 스톱워드(stop words)를 제거하고 스템밍 클래스(stemming classes)를 사용하여 질의어를 정규화 시킨다. 그 다음 질의어에 가장 적합한 개념망 노드를 규명하고 그 노드로부터 시작해서 개념망을 브라우징하며 정보를 검색할 수 있도록 한다. 만약 질의어와 관련된 노드가 개념망에 존재하지 않으면 전형적인 정보검색 방법에 의해 문서를 검색하고 (즉 텍스트 워드 검색) 검색된 문서와 그 문서들의 키워드를 이용하여 서브 개념망을 구성한다. 이 경우 네비게이션은 서브 개념망에서 이루어지게 된다.

제안된 시스템은 Java, JavaScript 및 Java Servlets을 이용하여 웹 서버(Tomcat) 상에 구현되었으며 300개의 개인 정보를 프로토타입 시스템에 등록하였다. 그림 4는 사용자가 최상위 레벨에 있는 개념들 중 “Ontologies (34)”를 선택하였을 경우나 혹은 질의어 “ontology”를 입력하였을 경우의 검색 결과를 보여준다. 개념들은 하이퍼링크로 표현되어 있으며 괄호 안에 있는 숫자는 각 개념 노드에 존재하는 문서의 수를 나타낸다. 제 3장에서 명시한 바와 같이 하나의 개념은 문서집합과 그 문서집합의 원소들과 관계가 있는 키워드집합의 쌍으로 이루어지며 따라서 엄밀히 말해, 하이퍼링크상에 있는 용어는 한 개념의 키



〈그림 4〉 개념망을 기반으로 한 브라우징 인터페이스

워드집합에 해당하며 문서집합은 하이퍼링크에 연결되어 있는 문서들에 해당한다. 텍스트 박스는 질의어를 입력하기 위한 인터페이스이며 원하는 질의어를 입력한 후 검색을 진행 할 수 있다. 링크로 연결된 알파벳 문자는 각각의 문자에 대응하는 키워드 리스트로부터 검색을 원하는 키워드를 선택할 수 있도록 지원한다.

그림 4에서 “Ontologies (34)”를 제외한 하이퍼링크로 연결된 개념들은 개념망에서 “ontology”의 직계 자식노드에 속해 있는 개념들이다. 개념 “ontology”에 대응하는 검색 결과는 화면 밑 부분에 나타나 있다. 지면 관계상 검색 결과의 일부분만을 보인다. 하이퍼링크로 연결되어 있는 문서 제목을 클릭하며 문서의 내용이 나타난다. 각 개념의 왼쪽에 위치한 폴더이미지(☐)를 클릭하면 각 개념에 대응하는 검색결과를 볼 수 있다. 사용자는 서브 개념들 중 하나를 클릭하여 검색을 진행하거나 혹은 새로운 키워드를 입력하여 검색을 새롭게 다시 시작할 수 있다. 디스플레이 공간을 간편하게 나타내기 위해 각각의 서브개념으로부터 현재 개념에 해당하는 단어를 삭제하였다. 즉, 서브개념 “Conferences (8)”은 “Ontologies, Conferences (8)”의 축소된 형태이다.

그림 5은 사용자가 그림 4에서 “Ontologies(34)”의 서브 개념 “Semantic Web (12)”을 선택하였을

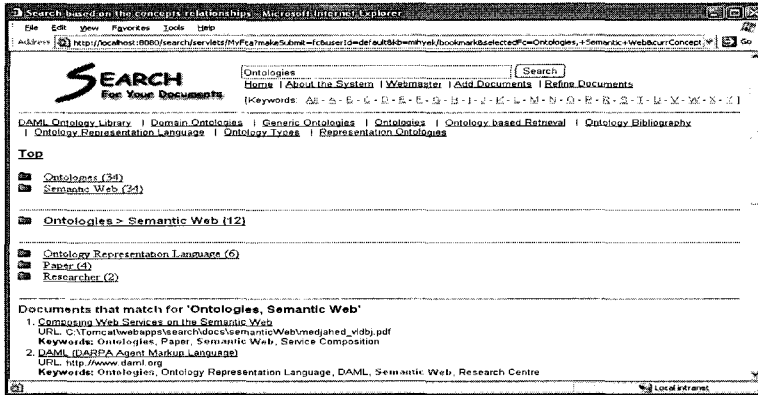
경우의 검색 결과를 보여주고 있다. 선택된 개념의 모든 직계 부모 노드와 자식 노드 개념들이 나타나 있다. 사용자의 이해를 돕기 위해 부모 노드는 빨강색으로, 현재 노드는 녹색으로, 자식 노드는 파랑 색으로 구분하여 표시하였다. 사용자는 좀 더 일반적인 문서 검색을 원할 경우는 부모 노드로, 좀 더 세부적인 문서 검색을 원할 경우는 자식 노드로 탐색을 진행할 수 있다. 사용자 질의어를 포함시키는 키워드의 메타데이터가 시스템에 존재하면 이러한 키워드들 또한 지원해 준다. 예를 들어, 화면 상단에 존재하는 키워드 “DMAL Ontology Library”, “Domain Ontologies” 등이 이러한 경우이다.

## 5. 실험결과 및 성능평가

### 5.1 실험결과

제안된 시스템과 기존 시스템과의 가장 큰 차이점은 자료저장에 있어 계층적 디렉터리 모델을 넘어 정보 사이의 연관 관계를 나타내는 개념망 구조 안에 문서들을 관리하여, 보다 효율적인 검색 성능을 제공하는데 있다.

문서들이 개념들 사이의 연관 관계에 의해 어떻게 색인화되어 관리되며 또한 색인화된 문서들



〈그림 5〉 개념망을 기반으로 한 브라우징 인터페이스의 주요 특징을 보여주는 예제

은 개념들 사이의 연관 관계에 의해 어떻게 검색되는지를 검토하기 위해 300개의 개인 정보들(자주 이용하는 웹 사이트, 연구 논문, 학회 정보 등)을 구현한 시스템에 등록하였다. 각각의 문서는 3내지 8개의 키워드 범위 내에서 평균 4개의 키워드로 주석이 되었으며 개념망은 노드 당 평균 2개의 부모 노드로 이루어진 280개의 노드들로 구성되었다. 루트에서 마지막 노드까지의 경로의 길이는 2에서 최대 5로 이루어졌다.

그림 6은 300개의 개인 정보를 이용하여 웹 상에서 구현한 개념망의 최상위 레벨에 있는 개념(예제1) 및 몇몇 개념들을 중심으로 한 개념망의 내부 구조를 평편한(flat) 형식으로 보여주고 있다. 괄호 안에 있는 숫자는 각 개념 노드에 존재하는 문서의 수를 나타낸다. 이 내부 구조를 통해 한 개의 부모 노드만을 허용하는 계층적 트리 구조와는 달리 개념망은 여러 개의 부모 노드 관계를 형성하는 망 구조에 기반하고 있으며 결과적으로 이러한 구조는 개념들 사이의 연관 관계에 의한 검색이 가능케 함을 알 수 있다. 예를 들어, 그림 5와 그림 6의 예제 4에서 볼 수 있듯이 사용자가 “Ontology” 개념으로부터 네비게이션을 시작하여 좀 더 세부적인 개념으로 “Semantic Web”을 선택하여 개념 “Ontology & Semantic Web”과 관련된 검색 결과를 확인한 후, 사용자는 부모 노드 중 하나인 개념 “Semantic Web”

을 브라우징할 수 있다.

즉, 사용자는 “ontology” 개념 하에 “semantic web”과 관련된 문서가 12개 존재함을 알 수 있고, 이 노드 “Ontologies > Semantic Web (12)”은 2개의 부모 노드를 가지고 있으며 이러한 망 구조로 표현된 뷰 (lattice view)을 통해 “semantic web”과 관련된 문서는 사실상 시스템에 34개가 있음을 알 수 있다. 사용자는 부모 개념 “Semantic Web (34)”을 선택하여 “semantic web” 개념의 관점에서부터 개념망을 관찰할 수 있게 된다. 그림 6의 예제 5에서도 같은 경우를 찾을 수 있으며 앞서 언급한 바와 같이 개념망은 노드 당 평균 2개의 부모 노드로 구성되었으며 이러한 결과는 개념들 사이의 연관 검색을 가능케 함을 증명한다고 볼 수 있다. 여기에서, “semantic web”는 “ontology”의 서브개념으로 분류될 수도 있지만 “semantic web” 개념의 관점에서 보면 “ontology”가 “semantic web”의 서브개념으로 분류될 수도 있음을 알 수 있다. 이처럼 FCA의 개념망은 문서들과 키워드 사이에 존재 가능한 모든 관계를 반영하여 개념들 사이에 일관성을 유지하며 개념들을 구조화 시킬 뿐만 아니라, 개념망 구조 안에 존재하는 모든 개념들 간에는 계층적 상속 관계가 유지되도록 한다. 물론, 계층적인 트리 구조를 사용하여도 모든 개념의 관점에서 클러스터를 체계화하여 분류할 수 있다. 하지만 수작업으로

모든 클러스터 사이의 연관 관계를 찾아 명시하고 문서가 추가됨에 따라 클러스터의 일관성을 유지한다는 것은 결코 쉬운 일이 아니다.

또한, 2장에서 언급한 바와 같이 계층적인 트리 구조를 이용하여 문서를 클러스터링 할 경우에는 보통, 문서를 대표할 수 있는 가장 적절한 클러스터 항목을 결정한 후 그 계층적인 경로 상에 문서를 색인한다. 그 결과, 정보 검색과 관련하여 카테고리 불일치 현상이 발생할 수 있다. 즉, 계층적인 분류 체계상에 있는 일반적인 용어 (general terms) 중에는 해당 문서와 아무 관련이 없는 클러스터 용어가 존재할 수 있다. 예를 들어, “Artificial Intelligence”와 “Databases” 양쪽에 클러스터된 “Data Mining”이라는 용어가 있다고 가정해 보자. 이 때 데이터 마이닝과 관련된 하나의 문서가 있다고 가정하면, 그 문서는 인공지능이나 데이터베이스 양쪽 혹은 어느 한쪽의 클러스터와 관련이 있을 수도 있지만, 때로는 그 어느 클러스터와도 무관할 수 있다. 이러한

경우에 이 문서는 부적절하게 클러스터될 수밖에 없다. 하지만, FCA는 문서와 키워드 사이에 함축되어 있는 모든 관계를 추출하여 계층적 상속 관계를 유지하는 개념망을 형성한다.

개념망을 기반으로 한 브라우징 구조의 또 하나의 주요 장점은 하부 노드로 내려가며 하나의 경로를 따라 네비게이션을 진행하다 만약 원하는 문서를 찾지 못하면, 단순히 처음부터 검색을 다시 시작하기 보다는 그 시점에서 개념 사이의 상호 연관 관계에 따라 또 다른 경로로 올라가며 검색을 진행할 수 있다. 계층적인 트리 구조에서는 각 단계마다 최선의 선택을 하며 하부 노드로 검색을 진행하다 만약에 관련된 문서를 찾지 못하면, 이미 각 의사결정 시점 때마다 최선의 결정을 내렸기 때문에 그 다음에 무엇을 해야 하는지 아는 것은 매우 어려운 일이다. 그러나 개념망은 한 개의 부모노드를 허용하는 계층적인 트리 구조와는 달리 여러 개의 부모노드를 허용함으로써 여러 경로를 가진다. 따라서 개념망 구조

<b>예제 1: Root (300)</b>		
Books (6)	KA2 (2)	Research Centre (10)
Browsing (4)	Knowledge Acquisition (28)	Researcher (21)
Classification Schemes (5)	Knowledge Management (11)	Search Engines (27)
Conferences (43)	Ontologies (34)	Semantic Search (3)
DAML (5)	Paper (13)	Semantic Web (34)
Formal Concept Analysis (30)	Projects (10)	Thesaurus (4)
Information Retrieval (14)	Publications (9)	UNSW (7)
<b>예제 2: Root (300) =&gt; Ontologies (34)</b>		
Conferences (8)	Ontology Types (5)	Researcher (6)
Knowledge Acquisition (8)	Publications (4)	Semantic Web (12)
Knowledge Management (5)	Research Centre (4)	
<b>예제 3: Root (300) =&gt; Semantic Web (34)</b>		
Community Portal (2)	Ontologies (12)	Research Centre (3)
Conferences (2)	Paper (8)	Workshop (3)
DAML(3)	Projects (8)	Symposium (1)
<b>예제 4: Root (300) =&gt; =&gt; Ontologies (34) =&gt; Semantic Web (12)</b>		
<b>Parent Concepts:</b>		
Ontologies (34)	Semantic Web (12)	
<b>Sub Concepts:</b>		
Ontology Representation Language (6)	Paper (4)	Researcher (2)
<b>예제 5: Root (300) =&gt; Semantic Web (34) =&gt; Paper (8)</b>		
<b>Parent Concepts:</b>		
Semantic Web (34)	Paper (13)	
<b>Sub Concepts:</b>		
Information Retrieval (1)	Ontologies (4)	Web Services (3)

<그림 6> 개념망의 최상위 레벨에 있는 개념 및 몇몇 개념들을 중심으로 한 개념망의 예제들

는 다른 토픽들 간의 연관 관계에 의해 이전에 미처 고려하지 못한 새로운 의사결정을 내릴 수 있도록 하며, 또 다른 경로를 통해 자유롭게 상·하로 이동하며 검색을 진행할 수 있도록 개방되어 있다. 사용자의 중요한 검색 행위(search behaviour) 중의 하나는 정확하게 무엇을 찾고자 알고 있는 경우조차도 질의어에 기반 한 검색 엔진을 사용하지 않고 화면에 나타나 있는 문맥상의 지식이나 정보를 이용하여 몇 번의 네비게이션을 수행한 후 원하는 정보를 검색한다는 것이다[20]. 즉, 질의어에 기반을 둔 검색 엔진은 대개는 오리엔티어링(orienting) 전략을 위한 일부로만 사용되어질 뿐 사용자의 구체적인 검색 요구는 문맥 정보를 통해서 명확하게 완성된다는 것이다. FCA의 개념망 안에 나타나 있는 토픽들 간의 연관 관계는 이러한 사용자 검색 행위를 충족시킬 수 있는 검색 도구로서의 역할을 충분히 수행하였다.

브라우징을 위한 FCA 개념망의 또 다른 특징은 개념망 내에 있는 한 정점에서 사용자 질의어에 대하여 모든 최소 확장(minimal enlargement)과 최소 변경(minimal refinement)을 내포한다는 점이다[21]. 즉 사용자 질의어에 대응하는 정점에서 간선으로 연결된 직계 상위(하위) 개념은 그 질의어에 대한 최소 확장(최소 변경) 정점에 해당한다. 예를 들어, 질의어 “Ontologies, Semantic Web”을 이용하여 정보를 검색하다 이 질의어와 관련이 있으면서 좀 더 세부적인 정보를 얻고자 할 경우 직계 하위에 있는 개념들이 바로 이 질의어를 최소로 확장 시킬 수 있는 개념들이나 것이다. 따라서 개념망 구조는 사용자가 질의어를 재정의하고자 할 때 도구 역할을 할 수 있다.

## 5.2 성능평가

제안된 시스템의 망 기반 브라우징 메커니즘의 검색 성능을 계층적 트리 구조에 기반 한 시스템의 성능과 비교하기 위하여 개인의 문서를 관리하기 위해 일반적으로 가장 많이 이용하고 있는 Win-

dows XP의 파일 시스템을 사용하였다. Windows 파일 시스템은 계층적인 트리 구조를 사용하는 대표적인 경우이며 폴더를 이용하여 문서를 분류하여 관리한다. 폴더의 이름은 분류체계 시스템의 클러스터에 대응된다고 볼 수 있으며 폴더의 계층적인 구조는 그래픽으로 표현된다. 검색 기능은 폴더의 이름과 파일의 이름에서, 혹은(그리고) 파일의 내용 중에서 해당 검색어를 포함하고 있는 문서를 찾아 표시한다. 실험에서는 폴더의 이름과 파일의 이름 그리고 파일의 내용에서 해당 질의어를 포함하고 있는 검색 결과를 사용하였다. 실험을 위해 프로토타입에 의해 생성된 개념망을 트리 구조로 변환하여 Windows 파일 시스템에 등록하였다. 개념망의 구조를 트리 구조로 변환할 때 개념망의 최상위 레벨에 있는 개념들에 대하여 레벨 3가지의 자식 개념을 사용하여 계층적인 분류체계를 구성하였다. 즉, 개념망의 최상위 레벨의 있는 개념은 파일 시스템의 최상위 폴더에, 자식 개념들은 각자 대응하는 최상위 폴더의 서브 폴더에 위치한다.

검색 성능 비교에 있어, 개인 정보관리자가 본인이 관리하고자 하는 문서에 대하여 적합한 개념을 주석할 수 있는 가장 합당한 에이전트라는 가정 하에, 개념망 내에서 정보관리자에 의해서 주석된 키워드들에 대한 검색 결과는 정확률(precision)과 재현율(recall)에 있어 완전하다고 가정하였다. 이러한 가정 하에 정보관리자가 문서를 주석하기 위해 사용한 236개의 키워드에 대하여 검색 성능을 계산하였다. 결과는 Windows 파일 시스템이 정확률과 재현율 둘 다에서 개념망의 검색 성능보다 낮은 결과를 보였다. 파일 시스템의 정확률과 재현율은 각각 0.62와 0.86으로 나타났다. 정확률에 있어 낮은 성능을 보인 것은 검색된 문서가 찾고자 하는 문서는 아니었지만 해당 문서의 내용 중에 검색어가 포함되어 있어 검색된 경우가 주 원인으로 작용하였다. 재현율에 있어서는 대체로 높은 성능을 보였지만 이러한 결과는 개념망에 근거하여 계층적인 분류체계를 구성하

였기 때문이라 여겨지며, 다소의 차이점을 보인 것은 개념망을 계층적인 분류체제로 변환 시 레벨 3이하는 무시하였기 때문이라 사료된다.

비록 이러한 성능 평가가 완전한 것은 아니지만 위의 실험 결과는 개인 정보를 관리하는데 있어 개념망 접근방법이 계층적 방법보다는 보다 우위에 있음을 뒷받침할 수 있다고 본다. 실험에서 관찰된 이를 뒷받침할 수 있는 또 하나는 계층적 분류 체계에서는 클러스터간의 연관성을 나타낼 수 있는 방법이 없기 때문에 동일한 문서를 중복하여 관리할 수밖에 없고, 예를 들어, *Ontology*, *Semantic Web* 과 관련된 문서가 있다고 가정할 때 그 문서는 “*Ontology => Semantic Web*” 과 “*Semantic Web => Ontology*” 양쪽 클러스터에서 각각 관리되며, 그 결과 해당 용어중의 하나로 검색을 하였을 경우 중복된 결과가 나타나는 것을 실험에서 관찰하였다.

## 6. 결 론

FCA는 지식표현과 처리를 위한 여러 분야에 활용될 수 있다[5]. FCA의 개념망은 또한 자료 분석과 정보검색을 위한 정보 과학, 소프트웨어 공학, 라이브러리, 의학 등 다양한 분야에 많은 응용 시스템을 가지고 있다. 지식표현의 경우 문맥 논리(contextual logic)에서 추론을 향상시키기 위한 방법으로 지식표현에 널리 사용되고 있는 의미 망과 논리를 기반으로 개념들 간의 관계로 구성된 개념그래프(Conceptual Graphs)와 통합하여 사용할 수도 있다[22].

본 논문에서는 조직이나 개인이 필요로 하는 문서를 웹을 기반으로 하여 자유롭게 관리하고 유지할 수 있는 더불어 관리된 문서를 쉽게 검색할 수 있는 정보관리시스템을 구축하기 위해 FCA 기법을 사용하는 것을 제시하고 개인이 자주 이용하는 웹 사이트 정보나 연구 논문 등의 자료에 대한 정보관리·검색 시스템을 구현하였다.

문서 정보는 FCA의 개념망으로 색인하여 관

리하며 개념망의 구조는 새로운 정보가 추가 변경되어 감에 따라 개념들 사이의 계층적 상속 관계를 일관성 있게 유지하며 자동적이고 점진적으로 발전 전개되어 간다. FCA의 개념망은 브라우징 메커니즘으로 사용하여 문서들과 문서 속성 사이에 존재 가능한 모든 관계를 반영한 실행 가능한 모든 네비게이션 경로를 지원함으로써 개념 사이의 연관 관계에 의한 검색을 지원할 뿐만 아니라 다양한 경로를 탐색하며 보다 더 정확한 정보를 얻을 수 있도록 지원한다. 일반적인 검색 엔진의 전형적인 방법인 질의 인터페이스를 통한 텍스트 워드 검색 또한 지원한다. FCA 접근방법의 가치는 무엇보다도 계층적 접근방법의 단점을 극복할 수 있는 망 기반 브라우징(lattice based browsing) 메커니즘에 있다. 본 논문에서는 이러한 브라우징 시스템의 장점에 대하여 실험 결과를 바탕으로 설명하였으며 문서를 검색하는데 있어 개념망을 기반으로 한 브라우징은 계층적인 구조를 기반으로 하는 시스템보다 더 나은 성능 평가를 나타냄을 보였다.

향후에는 텍스트 마이닝 기술을 이용한 보다 정확한 속성 추출을 통한 FCA 개념망 구축 및 상용화된 검색 엔진의 장점과 현재 지식관리 방법의 핵심 기술로 국내외적으로 크게 대두되고 있는 시맨틱 웹(Semantic Web)을 기반으로 한 의미 기반 온토로지 접근 방법과 접목하여 보다 유연하고 지능화된 시스템에 대한 연구를 수행할 계획이다.

## 참 고 문 헌

- [1] Ding, Y., Fensel, D., Klein, M., Omelayenko, B., “The Semantic Web: Yet Another Hip?”, *Data and Knowledge Engineering*, 41(3), pp. 205-227, 2002.
- [2] Clancey, W. J., “*Situated Cognition: On Human Knowledge and Computer Representation*”, USA, Cambridge University

- Press, 1997.
- [3] Peirce, Ch. S., "Collected Papers of Charles Standers Peirce", Cambridge, Harvard University Press, 1931.
- [4] 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.
- [5] Ganter, B., Wille, R., "Formal Concept Analysis: Mathematical Foundations", Heidelberg, Springer, 1999.
- [6] Zamir, O. and Etzioni, O., "Grouper: A Dynamic Clustering Interface to Web Search Results", *Computer Networks*, 31(11-16), pp. 1361-1374, 1999.
- [7] Marchionini, G., "Information seeking in electronic environments", New York, Cambridge University Press, 1995.
- [8] Furnas, G. W., Landauer, T. K., Gomez, L. M. and Dumais, S. T., "Statistical semantics: analysis of the potential performance of key-word information systems", *Bell System Technical Journal*, 62:1753-1806, 1983.
- [9] Maedche, A. and Staab, S., "Mining Ontologies from Text", 12th European Conference on Knowledge Acquisition and Knowledge Management (EKAW 2000), Springer, pp. 189-202., 2000.
- [10] Handschuh, S., Staab, S., Ciravegna, F., "S-CREAM - Semi-automatic CREATION of Metadata", 13th International Conference on Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web (EKAW 2002), Springer, pp. 358-372, 2002.
- [11] Staab, S., Angele, J., Decker, S., Erdmann, M., Hotho, A., Maedche, A., Schnurr, H. P., Studer, R., Sure, Y., "Semantic Community Web Portals", Proceedings of the 9th International World Wide Web Conference, pp. 474-491, 2000.
- [12] Davies, J., Duke, A., Sure, Y., "Onto-Share", Proceedings of the Second International Conference on Knowledge Capture (K-CAP 2003), New York, ACM, pp. 20-27, 2003.
- [13] Quan, D., Karger, D.R., "How to Make a Semantic Web Browser", Proceedings of the 13<sup>th</sup> International Conference on World Wide Web, pp. 255-265, 2004.
- [14] Guha, D., McCool, R., Miller, E., "Semantic Search", Proceedings of the 12<sup>th</sup> International Conference on World Wide Web, pp. 700-709, 2003.
- [15] Godin, R., Missaoui, R. and April, A., "Experimental Comparison of Navigation in a Galois Lattice with Conventional Information Retrieval Methods", *International Journal of Man-Machine Studies*, 38:747-767, 1993.
- [16] Carpineto, C. and Romano, G., "Information retrieval through hybrid navigation of lattice representations", *International Journal of Human-Computer Studies*, 45:553-578, 1996.
- [17] Cole, R. and Eklund, P., "Browsing Semi-structured Web texts using Formal Concept Analysis", Proceedings of the 9th International Conference on Conceptual Structures (ICCS 2001), Stanford, Springer, pp. 290-303, 2001.
- [18] Kim, M., Compton, P., "Evolutionary Document Management and Retrieval for Specialised Domains on the Web", *International journal of human computer*

- studies, 60(2), pp. 201-241, 2004.
- [19] Stumme, G., "Hierarchies of Conceptual Scales", 12th Banff Knowledge Acquisition, Modelling and Management (KAW'99), Banff, Canada, SRDG Publication, 5.5, pp. 1-18, 1999.
- [20] Teevan, J., Alvarado, C., Ackerman, M.S., Karger, D.R., "The Perfect Search Engine is not Enough: A Study of Orienteering Behavior in Directed Search", Proceedings of the SIGCHI conference on Human Factors in Computing Systems, ACM, pp. 415-422, 2004.
- [21] Godin, R., Missaoui, R. and Alaoui, H., "Incremental concept formulation algorithms based on Galois (concept) lattices", Computational Intelligence, 11(2), pp. 246-267, 1995.
- [22] Wille, R., "Conceptual Graphs and Formal Concept Analysis", Proceedings of the 5th International Conference on Conceptual Structures (ICCS'97), Springer, pp. 290-303, 1997.

## ● 저 자 소 개 ●



### 김 미 혜 (Kim, Mihye)

1984년 전북대학교 전산통계학과 졸업(학사)  
 1999년 New South Wales 대학교 대학원 컴퓨터공학과 졸업(석사)  
 2003년 New South Wales 대학교 대학원 컴퓨터공학과 졸업(박사)  
 2004년 9월~현재 대구가톨릭대학 컴퓨터교육과 교수  
 관심분야 : 지식 관리 및 검색, 온토로지, 컴퓨터교육 etc.  
 E-mail : mihyekim@cu.ac.kr