

# 구조적 분류 기법을 기반으로 한 정보 시각화 시스템 설계 및 구현

김 영 란<sup>†</sup> · 구 연 설<sup>††</sup>

## 요 약

대량의 정보를 대상으로 하는 웹 정보 검색 기술은 정보를 수집하는 방법과 브라우저의 시각화 인터페이스 기법에 대해 활발히 연구가 진행된 반면에, 수집된 정보를 효율적으로 분류하는 구조적 모델링에 관한 연구는 미흡한 실정이다. 이 논문에서는 정보의 구조적 모델을 기반으로 하여 정보의 특성을 시각적인 기법으로 제공함으로써 사용자의 검색 환경을 효율적으로 운용하는데 목적이 있다. 이를 위해 정보가 갖는 특성을 기반으로 하여 정보를 Facet 단위로 분류하는 구조적 방법을 제안하고, 정보간의 관련성을 정의한다. 또한, 각 Facet 단위의 정보가 갖는 특성에 대한 분석 결과를 다중 뷰를 통해 시각화한다. 따라서, 사용자는 검색 대상 범위의 정보에 대한 체계적인 구조 및 시각화된 정보를 인지한 후, 좀 더 쉽게 검색을 수행할 수 있다. 중심어 매칭 방식에 의한 검색 결과 목록을 제시하는 기존의 검색 기법이 정보를 찾는데 중점을 둔 반면에, 이 논문은 저장된 정보가 무엇이 있는지를 구조적으로 표현하여 사용자의 이해도를 향상시킴으로써 사용자의 관점에 중점을 둔다.

## Design and Implementation of an Information Visualization System based on Structured Classification Technique

Young-Ran Kim<sup>†</sup> · Yeon-Seol Koo<sup>††</sup>

## ABSTRACT

While the method of information collection and visual interface technique have been researched actively on web information retrieval, a study on structured modeling for effective classification of a wide collective information leaves to be desired. In this paper, we represent information feature based on structured information model. It aims at carrying out effectively the user's retrieval environment through visualization technique with analyzing the information feature. We propose a information classification method using Facet units and we construct the object model, table model, SQL code to define the relation of the information, and represent the information feature based on a wide range of views. After users gain a better global understanding of the information feature, retrieve more easily through their information. Conventional information retrieval is user-oriented to be what user want, but proposed technique is data-oriented which helps users to understand what exist in database by showing information feature.

### 1. 서 론

일반적으로 인터넷 사용자는 검색 도구의 기능을 배

우고 생각해가며 질의어를 만들기보다는 원하는 정보를 시각화된 인터페이스를 통해 보다 편리하게 찾기를 원한다. 사용자와 정보간의 효율적인 인터페이스를 제공하는 기법들은 정보처리 기법, 재사용 기법, 브라우저 기법, 검색 기법 등을 고려하는 방안으로 변화하였다[1]. 따라서, 정보 검색 시스템은 단순히 검색 기능뿐만 아니

<sup>†</sup> 정 회 원 : 충청대학 컴퓨터학부 교수

<sup>††</sup> 정 회 원 : 충북대학교 컴퓨터과학과 교수

논문접수 : 1999년 4월 26일, 심사완료 : 1999년 10월 20일

라 정보 라우팅 및 여과, 브라우징, 정보 시각화 등의 다양한 서비스를 제공해야 한다[2, 3]. Hurse에 의하면, 디지털 도서관을 개발하는 주목적 중에 하나는 사용자 중심으로 정보를 검색하고 사용할 수 있도록 기회를 제공하는 것이다. 이를 위해 각 응용 분야에서 적절하게 정보 객체들을 분류하고, 분류된 객체들을 효과적으로 화면에 시각화하여야 한다[4].

기존의 정보 검색 기법 및 활용 도구를 이용한 검색 환경은 대규모의 정보 처리 작업과 정보 검색의 효율성 관점에서 다음과 같은 문제점을 갖는다. 첫째, 주어진 중심어에 대해 너무 많은 결과 항목을 생성한다. 둘째, 불 대수 조합에 의한 질의는 사용자가 원하는 정보 집합을 직접적으로 기술하도록 제한한다. 셋째, 반환된 정보 목록의 순서는 정보간의 관련성을 반영하지 못한다[5]. 이로 인해, 사용자는 관련성 없이 나열된 수많은 결과 목록을 향해함으로써 시간 낭비를 초래하고, 정보에 대한 전반적인 구조에 대한 개념 없이 검색을 수행하다가 방향성을 상실함으로써 찾고자 하는 정보를 검색하지 못하게 된다.

정보의 내용 검색에 중점을 둔 기존의 정보 검색 기술은 정보의 구조적 모델을 기반으로 한 검색을 고려하지 않고, 대량의 자료를 효과적으로 분류하여 저장하고 관리하는 기술은 초기 단계에 있는 실정이다[2]. 정보에 포함된 내용과 정보간의 관련성 구조에 대한 개요 없이 효율적인 검색을 수행하는 것은 어려운 일이다. 기존의 문제점을 해결하기 위해, 이 논문에서는 수많은 정보의 공통적인 특성과 관련성을 효율적으로 시각화하는 것을 목적으로 한다. 연구 방법으로는 정보가 갖는 공통적인 특성과 이들간의 관련성을 체계적으로 나타낼 수 있는 구조적 정보 분류 기법(Structured Information Classification Technique)을 제안한다. 또한, 제안된 기법과 브라우징 기법을 통합함으로써, 정보에 대한 사용자의 이해도를 향상시키고 보다 쉽게 정보를 향해할 수 있는 정보 시각화 시스템을 설계 및 구현한다.

이 논문의 구성은 다음과 같다. 2장에서는 웹 검색 시스템의 인터페이스 연구 동향 및 구성요소를 기반으로 한 분류 기법을 기술하고, 3장에서는 정보가 갖는 특성을 시각화하기 위한 구조적 정보 분류 방안을 제안하고 시스템의 객체 지향 분석 모델을 기술한다. 4장에서는 정보 시각화 시스템에 대한 기능과 정보 표현을 위한 관계 정의 및 색인 구조를 설계하고 구현한

다. 5장에서는 이 논문의 결론 및 향후 연구과제를 기술한다.

## 2. 관련연구

Infogrid, Butterfly, WebBook 등과 같은 정보 시각화 시스템들은 정보 시각화 프로젝트(Information Visualizer Project)를 수행하는 중에 Xerox PARC에 의해 생성되었다. Infogrid 시스템은 문서 영역에 대한 시각화와 관리 측면을 강조하기 위해 영역에 대한 접근 용이성(ready-to-hand) 개념을 이용함으로써 시각화 도구로서의 유용성을 제공한다[7]. Butterfly View는 서비스에 대한 비동기적인 요청(request)을 제어하고 3차원 화면을 통해 도서 목록을 효과적으로 제시해준다[8]. WebBook은 사용자가 웹 활동에 대해 바인딩(binding) 작업을 통해서 조직적으로 체계화할 수 있도록 구성되어 있다[9]. 일본의 IBM 연구소에서는 전자 도서관의 시각화를 지원하기 위한 방안으로 다중 뷰 기법을 적용하여, 신문 기사 정보로부터 사용자가 원하는 정보를 시각적으로 검색 가능하도록 하였다[10]. 이와 같은 연구는 정보 시각화 방안에 중점을 둔 검색 시스템으로서, 정보의 체계적인 분류 방안보다는 검색 기법과 브라우징 기법을 통합하는데 치중한 연구이다. 따라서, 사용자가 쉽게 향해할 수 있도록 환경을 제공해 주는데 반해서, 정보의 개요를 파악할 수 있는 공통적인 특성과 관련성을 시각화하지 못하는 단점을 갖고 있다.

정보 검색은 보다 쉽게 적합한 정보를 찾는 것이고, 이러한 측면에서 분류 방법은 매우 중요하다. 잘 정의된 분류 구조는 효율적인 검색 시스템을 설계하는데 필수적이다. 특히, 웹 정보를 대상으로 한 인터페이스 개발은 정보의 특성이 반영된 정보 공간에 대한 체계적인 정보 분류 기법이 요구된다. 정보 모델의 기초 이론으로 가장 많이 이용된 것은 그래프 모델이다. 그 외에 추상 자료형(Abstract Data Type), 네트워크 모델, ER(Entity-Relation) 모델, 의미망 모델, 프레임 이론 등이 있다. 이러한 모델들은 정보 표현의 제약, 정보 특성 반영 부족 등의 문제점을 갖고 있다. 이와 같은 문제점을 해결하고 대규모의 정보 표현에 적합한 객체 지향 데이터 모델은 비교적 유연한 정보 표현이 가능하지만, 정보의 특성을 충분히 반영하지 못하는 단점을 갖는다. 구성요소를 기반으로 한 정보 분류 구

조로는 계층구조[12]와 중심어를 기반으로 한 구조[13]가 있다. 계층구조는 종속관계와 포함관계의 원리를 기반으로 하고 나무구조와 격자(lattice) 형태가 있다. 이 구조는 대량의 정보가 갖는 복잡한 관련성을 표현하는데 어렵다. 중심어를 기반으로 한 구조는 정보를 계층적으로 정확하게 분류하기에 곤란한 경우를 고려한 방법으로서, 정보간의 관련성에 의한 분류가 아니라 공통적인 특성에 의한 분류이다. 이 구조는 정보의 수가 증가하는 경우에, 하나의 클래스에 너무 많은 정보를 그룹화 하는 단점이 있다. 다른 분류 방법으로는 네트워크 구조가 있다[14]. 모든 정보를 포인트에 대한 네트워크를 통해 연결한다. 이 형태는 여러 개의 나무구조를 연결할 수 있기 때문에 복합(compound) 구조를 쉽게 구성할 수 있다. 반면에, 정보의 위치를 정하기 위해 긴 행로를 초래하는 오버헤드가 발생한다.

이 논문에서는 웹 정보의 특성이 계층적으로 표현되기 어렵다는 점과 정보간의 효율적인 관련성 시각화를 고려하여, 중심어를 기반으로 한 구조적 정보 분류 기법을 연구한다.

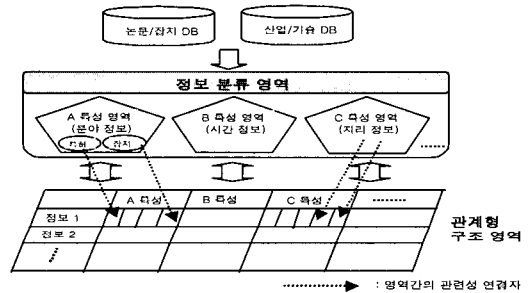
### 3. 정보 시각화 시스템 분석

#### 3.1 구조적 분류 기법

정보간의 관련성을 가장 잘 표현하는 관계형 모델을 기반으로 하고, 정보의 공통적인 특성을 효율적으로 구성하기 위한 구조적 정보 분류 기법을 제안한다. 제안된 기법에서는 정보의 세부적인 특성에 대한 분류를 배제하기 위해, 정보가 갖는 공통적인 특성을 나타내는 기본 단위로 Facet[13]을 사용한다. 구조적 정보 분류 방법은 (그림 1)과 같이 정보가 갖는 공통적인 특성을 표현하는 정보 분류 영역과 정보 그룹간의 관련성을 정의한 관계형 구조 영역으로 구성된다. 정보 분류 영역에서는 정보의 특성을 기능적 Facet과 환경적 Facet으로 분류하고, 기능적 Facet으로는 분야 정보를, 환경적 Facet으로는 시간 정보, 지리 정보로 구성한다. 분류된 자료들은 분야에 따른 테이블, 시간에 따른 테이블, 지리적인 위치에 따른 테이블, 그리고 이들 테이블간의 관련성을 정의하기 위한 동적 테이블로 구성한다. 동적 테이블은 사용자의 질의에 따라 부가적으로 생성되는 관계 정의의 테이블이다.

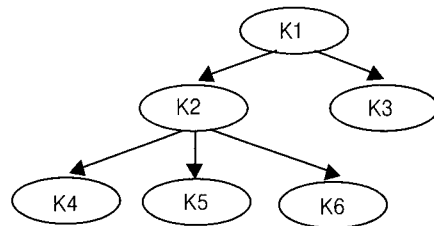
정보 분류 영역은 각 특성에 대한 영역(Domain)들의 집합으로 구성되고, 이들 영역간에는 독립적인 관

계가 존재한다. 예를 들면, A특성 영역과 B특성 영역간에는  $Dom(A특성) \cap Dom(B특성) = \emptyset$  이다. 또한, 정보 분류 영역에 포함된 임의의 정보에 대해,  $A특성(정보) \subseteq Dom(A특성)$  이다. 예를 들면, 잡지 정보가 갖는 분야별 특성은 분야 정보 영역의 부분 집합이고, 분야 정보 영역과 시간 정보 영역은 분명하게 구분된다. 데이터베이스에 저장된 정보 분류 영역에 대한 집합  $\{A특성, B특성, C특성, \dots\}$ 은 관계형 구조 영역을 정의하는 항목으로 사용된다. 이 구조는 계층적 구조와 관계형 구조의 통합형태로서, 관계형 모델의 장점을 이용하여 브라우징을 쉽게 할 수 있도록 지원할 뿐만 아니라, 정보의 특성을 변경하거나 추가 및 삭제하기가 용이하다.



(그림 1) 구조적 정보 분류 방법

정보간의 관련성은 추출된 정보 영역간의 종속 관계를 나타내는 것이고, 이러한 종속 관계는 정보 집합에 대한 사용자의 이해도를 높이기 위한 것이다. 일반적으로 웹 정보는 독립적인 특성을 갖는 정보 보다 관련성 있는 정보가 많다. 따라서, 효율적인 검색을 위해서는 적은 경우의 독립적인 관계를 식별하여 이를 배제하는 것이 바람직하다. (그림 2)는 정보 분류 영역에 존재하는 임의의 K특성에 포함된 속성 항목간의 관련



독립적인 특성을 갖는 집합: {K2, K3}, {K3, K4, K5},

(그림 2) 분류 영역의 관련성 그래프

성 그래프이고, 임의의 두 노드간에 경로가 존재하지 않으므로써 정보간에 독립적인 특성을 갖는 집합으로 는 {K2, K3}과 {K3, K4, K5, K6}이다.

3.2 경로 탐색을 위한 표현방법

제안된 구조적 정보 분류 구조에 대한 경로 탐색은 각 정보 영역에 포함된 항목간의 관련성을 표현한 그래프와 매트릭스를 기반으로 하여 수행된다. 먼저, 각 노드에 대한 종속된 특성을 파악하기 위해 노드의 레벨(level)을 구성한 후, 각 노드의 부분집합을 통해 독립적인 특성을 갖는 정보를 파악함으로써, 정보간의 관련성을 표현한 매트릭스로 부터 사용자가 원하는 정보를 탐색한다. 1단계에서 각 노드에 대한 종속 집합과 레벨을 구성하는 절차는 다음과 같다.

- a. 모든 노드의 수를 파악
- b. 현재 노드의 레벨을 1로 초기화
- c. While 레벨화된 노드의 수 ≠ 총 노드의 수

For 모든 노드에 대하여

If 현재 노드가 레벨화되어 있지 않고

노드의 input branch의 노드가 레벨화되어 있다면

Then 노드의 레벨을 현재의 레벨로 결정

Next

2단계에서 매트릭스 표현을 위한 독립적인 집합을 구성하는 절차는 다음과 같다. N개의 노드를 갖는 트리 구조에 대하여, 각 노드는 {k, k+1, ..., N-1}의 부분집합을 갖도록 레벨화가 되어 있다고 가정하자.

- a. root 노드에 대한 레벨화 집합 = {q|q = k 또는 q < n, q와 k은 독립적임}
- b. 만약,  $N = \{n_1, n_2, \dots, n_k\}$ 이고 i번째 노드가 레벨화되어 있는 경우에 대해,

If  $i \geq k$  Then 자식 노드를 갖지 않음

$D = \{n_k | k > i\}$  이고  $n_k$ 와  $n_i$ 는 독립적임

If  $i < k$  이고  $D = \{ \}$  Then

집합 N과 i+1번째 노드를 포함하여 레벨화된 자식 노드를 갖음

If  $D \neq \{ \}$  Then 두 개의 자식 노드를 갖음

하나의 자식 노드는 i번째 노드와  $\{n_1, n_2, \dots, n_i\}$ 을 포함하여 레벨화되고, 다른 자식 노드는 i+1번째 노드와  $\{n_1, \dots, n_i\} - D$ 를 포함하여 레벨화되어 있음

(그림 2)의 관련성 그래프에 대해, 1단계 과정에 의해 생성된 결과로는 K1노드의 레벨은 1이고 K2, K3노드의 레벨은 2이고 K4, K5, K6노드의 레벨은 3이다. 2단계 과정에 의해 독립적인 특성을 갖는 집합으로 추출된 경우는 {K2, K3}와 {K3, K4, K5, K6}이다. 왜냐하면, 레벨 1의 노드 집합 {K1}과 레벨 2의 노드 집합 {K2, K3}의 결합에 의해 노드 집합 {K1, K2, K3}이 생성되지만, K1과 K2는 종속적이다. 레벨 2의 노드 집합 {K2, K3}와 레벨 3의 노드 집합 {K4, K5, K6}의 결합에 의해 생성되는 노드 집합은 {K2, K4, K5, K6}와 {K3, K4, K5, K6}이 생성되지만, K2와 K4는 종속적이다. 따라서, 독립적인 특성을 갖는 모든 집합은 각 레벨에 대한 노드 집합 {K2, K3}, {K4, K5, K6}와 결합에 의해 생성된 노드 집합 {K3, K4, K5, K6}이 있다. 이 중에서 {K4, K5, K6}는 {K3, K4, K5, K6}의 부분집합이다. 독립적인 관련성을 갖는 노드 집합 {K2, K3}와 {K3, K4, K5, K6}을 매트릭스로 표현 하면 (그림 3)과 같다.

	K1	K2	K3	K4	K5	K6
K1	1	1	1	1	1	1
K2	1	1	0	1	1	1
K3	1	1	1	0	0	0
K4	1	1	1	1	0	0
K5	1	1	1	1	1	0
K6	1	1	1	1	1	1

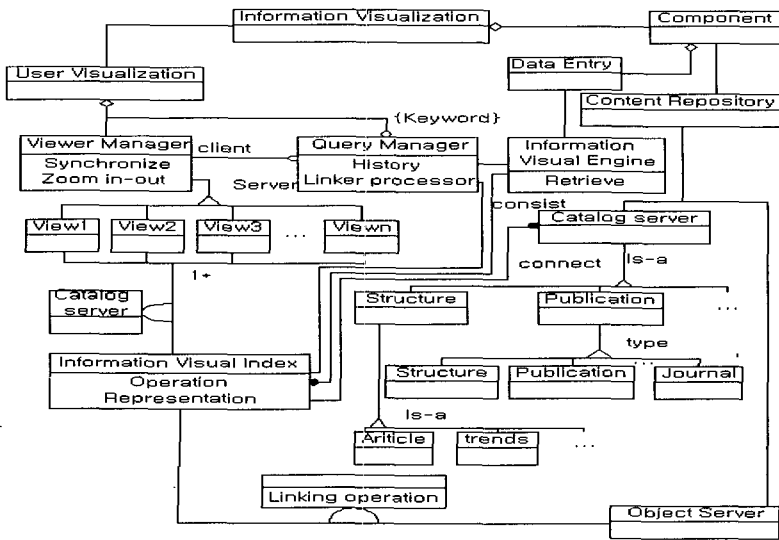
(1 : 종속관계, 0 : 독립관계)

독립적 관계 구성요소 : K2->K3, K3->K4, K3->K5, K3->K6, K4->K5, K4->K6, K5->K6

(그림 3) 관련성에 대한 매트릭스

3.3 시스템 분석을 위한 객체 지향 모델링

사용자는 정보의 영역을 표현한 다중 뷰 환경을 통해 자료를 검색하고, 각 뷰는 정보 영역의 특성을 시각적으로 표현한다. 여기에서는 시스템을 구성하는 객체 모델과 시스템의 기능을 분석하기 위한 기능적 모델을 기술한다. 객체 모델은 객체나 클래스들의 정적인 자료 구조와 그들 사이의 관계를 객체 다이어그램 도표를 이용하여 그래픽하게 표현한다. 객체 모델의 상속성, 다중성, 집합 관계 등을 기반으로 하여 관련성을 효율적으로 표현한 객체 모델은 (그림 4)와 같다. View manager와 Query manager간의 관계(association), 색인을 기반으로 하여 다중 뷰를 설정하도록 하는 연계 속성(link attribute), 그리고 정보 저장소에 저장된 정보 유형과 인터페이스를 연결시켜 주는 다중 관

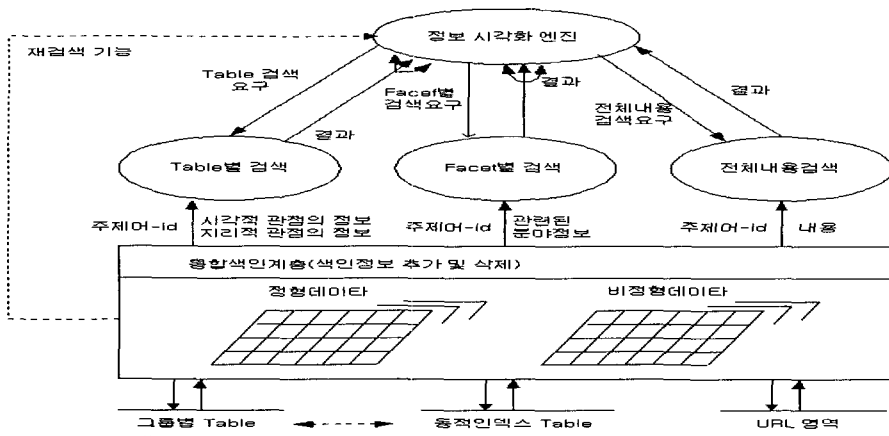


(그림 4) 객체 모델

계(multiple of association)를 표현한다. User interface 객체와 Retrieval system component 객체는 직접적인 인스턴스를 갖지 않는 추상화 객체이다. View manager 객체는 질의 영역에 따라 변하는 유동적인  $n$ 개의 뷰 집합(aggregation)으로 구성되며, 뷰 간의 상호 관련성을 파악하여 관련된 뷰로 이동할 수 있다. 이를 위해 View manager 객체는 동기화(synchronize) 메소드를 제공한다. 또한, View manager 객체와 Query manager 객체간에는 Client-server 관계가 존재하며, Query manager 객체에는 질의가 입력되지 않는 경우가 있을 수 있다. Information visual index 객체는 정보 저장소

에 저장된 자료의 특성을 추출하여 표현한 것으로서, 뷰와 mapping operation 연계 속성을 갖는다. 또한, 저장소에 저장된 Object server 객체와는 linking operation 연계 속성을 가지며, 이는 객체간의 관계성을 기반으로 하여 수행된다.

기능적 모델은 시스템 내부에서 자료 값이 변형되는 절차를 기술하는 것으로서, 연산과 제약 조건들의 의미를 기술하는 자료 흐름도로 표현한다. (그림 5)에서 보는 바와 같이 정보 시각화 시스템에 주제어-id 자료가 입력되면, Facet별 검색에 의한 분야 정보와 테이블별 검색에 의한 시각적 정보를 출력한다. 이때, 분야



(그림 5) 기능적 모델

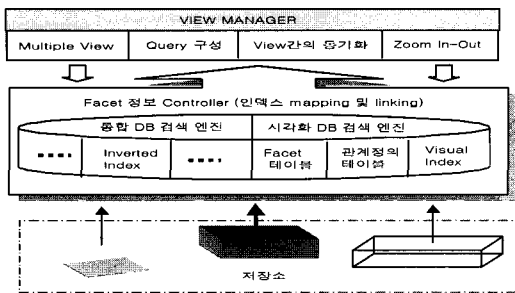
정보와 시각적 정보는 통합 색인 계층에 있는 형식화된 공통 항목에 대한 색인 정보와 임의의 정보만이 갖고 있는 공통 항목이외의 자료에 대한 관련성에 의해 구성된다.

#### 4. 시스템 설계 및 구현

이 논문은 Windows NT 환경에서 구현하였으며, 구현에 이용한 언어는 Delphi 4.0이다. 검색에 사용할 데이터베이스는 Access97을 이용하여 논문/잡지DB와 산업/기술DB의 레코드로 구성하였다.

##### 4.1 시스템 구성

각 뷰를 구성하는 정보의 특성은 통합 색인 데이터베이스(Total Index DB)와 시각화 색인 데이터베이스(Visual Index DB)에 저장되어 있다. 통합 색인 구조는 전체 자료에 대한 공통적인 항목의 특성을 표현하고 시각화 색인 구조는 각 뷰를 통해 나타나는 정보 영역과 이들간의 관련성을 정의한 관계 정의 테이블 및 동적 테이블을 표현한다. 정보 시각화 시스템의 구조도는 (그림 6)과 같다.



(그림 6) 시스템 구조도

##### 4.2 모듈 기능

###### 4.2.1 View manager

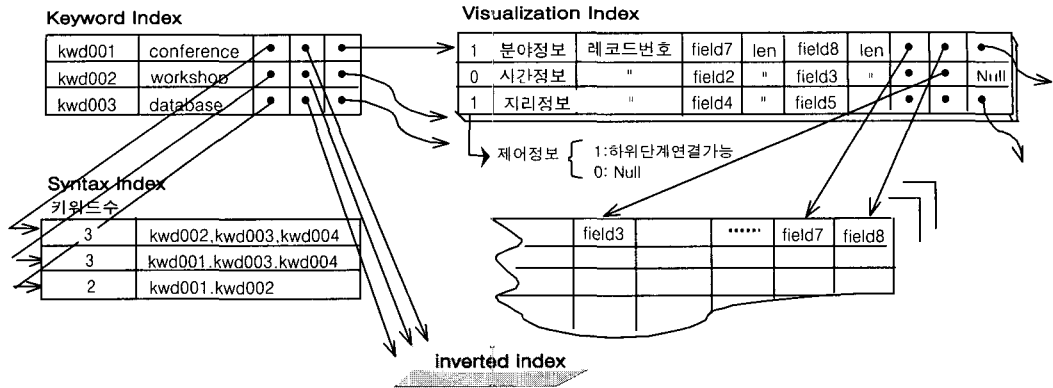
정보가 갖는 특성을 시각적으로 표현하는 View manager는 그래픽 객체의 집합으로 구성되고, 뷰간의 동기화뿐만 아니라 특정 영역에 대한 확대 기능을 지원한다. 다중 뷰로는 시간의 개념에 따라 발생된 정보의 수를 막대 그래프로 표현하는 시간 정보 뷰, 지역별 정보의 분포도를 지도 차트에 색상별로 나타내는 지역 정보 뷰, 정보의 분야별 분류에 따른 비율을 원형 그

래프로 나타내는 분야정보 뷰가 있다. 다중 뷰간에는 동기화가 가능하기 때문에, 사용자가 뷰간의 항해를 통해 점차적으로 원하는 정보에 접근하도록 지원한다.

###### 4.2.2 정보 관리자

제안된 구조는 브라우저를 보다 더 쉽게 하고 정보간의 관련성을 정의한 동적 테이블의 내용을 결합시킨다. 논문/잡지 데이터베이스와 산업/기술 데이터베이스의 공통 항목을 기반으로 하여 분야 정보, 시간 정보, 지리 정보로 분류한 후, 각 영역을 구성하는 정보들의 논리적인 표현으로 관계형 테이블 구조를 표현한다. 효율적인 정보 시각화를 위해 구조적 정보 분류 기법을 기반으로 하여, 공통 항목 이외의 특성과 관련성에 대한 색인 구조는 (그림 7)과 같다. 정보가 갖는 형식화된 공통 특성에 대한 색인 테이블을 구성하기 위해 레코드 번호, 검색 횟수, 로딩 일자, 국가명, 잡지명, 제목, 주제어, 요약 등 8개의 항목을 구성하였다. 또한, 다중 뷰를 구성하는데 사용되는 시간적 정보 및 정보의 영역을 표현하는 뷰 간의 관련성에 의한 이동이 가능하도록, 동기화를 지원하기 위한 동적 테이블을 구성한다. 색인 파일은 저장된 정보의 주제어와 정보 특성을 나타내는 레코드 및 레코드와 관련된 위치 정보인 포인터를 갖는다.

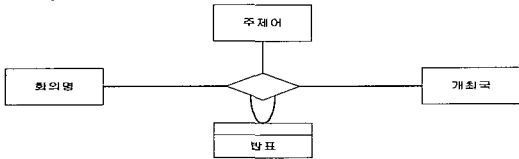
뷰 간의 이동이 가능하도록 하기 위해 정보 영역간의 관련성을 네트워크 구조로 표현한다. 이 구조는 정보들의 관련성을 포인터들로 이루어진 네트워크 구조에 의해 연결시키거나, 또는 3~4개의 계층 구조를 연결시킨다. 연결된 정보에 대해서는 다중 뷰 간의 관련성을 위해 테이블 정의 및 SQL 코드를 작성한다. 산업/기술 데이터베이스의 색인 항목으로는 주제어이고, 논문/잡지 데이터베이스의 학술 회의/세미나 테이블의 색인 항목으로는 회의명, 개최국, 주제어로 구성한다. 뷰 간의 연동이 가능하도록 하기 위해, 이러한 색인 항목간의 관계를 OMT 방법론[15]의 ternary association으로 정의한다. 다음은 주제어, 회의명, 개최국 간의 관련성을 객체 모델, 테이블 모델, SQL(Structured Query Language) 코드로 나타낸 것이다. 각기 독립적으로 존재하는 색인 항목에 대해, 관련성을 정의한 객체 모델은 주제어-ID에 대해 발표된 회의명과 개최국 간의 관련성을 정의함으로써, 다중 뷰를 구성하는 기본 요소를 설정하게 된다. 이러한 관계정의에 의해 뷰간의 동기화를 지원한다.



- 메타정보 스키마=<DB명, 테이블명, 메타순서번호, 메타이름, ...>
- Field명=<레코드번호, 검색횟수, 로딩일자, 국가명, 잡지명, 제목, 주제어, 요약>
- 레코드번호=그룹번호(2자리)+DB번호(2자리)+레코드순서(6자리)

(그림 7) 색인 구조

**Object Model**



Candidate key : (회의명-ID, 주제어-ID, 개최국-ID)

**Ternary Table Model**

속 성	Null?	Domain
회의명-ID	N	ID
주제어-ID	N	ID
개최국-ID	N	ID
발표	Y	Sessions

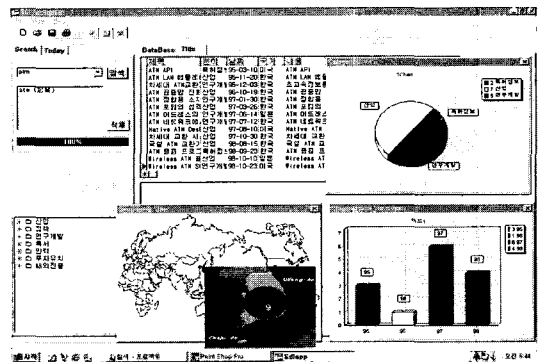
**SQL Code**

```

CREATE TABLE 주제어
(주제어-ID ID not null,
 and other attributes ... ,
 PRIMARY KEY(주제어-ID));
* 회의명과 개최국에 대한 테이블 구성
CREATE TABLE 주제어-회의명-개최국
(주제어-ID ID not null,
 회의명-ID ID not null,
 개최국-ID ID not null,
 발표 Boolean)
PRIMARY KEY (주제어-ID, 회의명-ID, 개최국-ID)
FOREIGN KEY (주제어-ID) REFERENCES 주제어.
FOREIGN KEY (회의명-ID) REFERENCES 회의명.
FOREIGN KEY (개최국-ID) REFERENCES 개최국;
CREATE SECONDARY INDEX 주제어-색인명
ON 주제어(주제어명)
    
```

**4.3 결과 화면**

사용자는 원하는 정보를 검색하기 위해 쉽고 편리하게 질의를 할 수 있다. 또한 질의 결과는 정보의 속성에 따라 각 뷰에 시각적으로 표현된다. 시각화 검색 시스템은 중심어를 이용한 검색 시스템과 질의 결과를 그래프 형식으로 나타내는 다중 뷰로 구성하였다. (그림 8)은 질의어 "ATM"에 대해 검색된 정보를 3개의 뷰 특성에 따라 시각적으로 표현한 화면이다. 각 뷰는 서로 동기화가 이루어지며, 뷰간의 항해를 통하여 점차적으로 원하는 정보에 접근할 수 있다. 예를들면, 입력된 질의 항목에 관한 정보를 [Title] 탭에서 목록으로 제시하고, 이들 정보가 갖는 특성을 지역정보 뷰, 분포도 정보 뷰, 비율정보 뷰를 통해 시각화한다. 또한, 지역 정보 뷰에서는 확대 기능을 지원해준다. 사용자가 연



(그림 8) 정보 시각화 시스템 화면

도에 따른 정보 분포도를 시각화한 뷰에서, 특정 막대 그래프를 선택하면 뷰간의 동기화에 의해, 해당 연도에 관련된 부분합 정보를 시각화한다.

## 5. 평가 및 결론

웹 정보 검색은 보다 쉽게 적합한 정보를 찾는 것이 중요하고, 이러한 문제를 해결하기 위한 측면에서 분류 방법은 매우 중요하다. 검색 방법은 분류 방법에 의해 영향을 받기 때문에, 잘 정의된 분류 구조는 효율적인 검색 시스템을 설계하는데 필수적이다. 이 논문에서는 정보 영역과 관련성을 기반으로 한 구조적 정보 분류 기법을 제안하였고, 각 영역에 존재하는 관련성에 대한 색인구조를 기술하였다. 이와 같은 정보 분류 구조를 기반으로 한 시각화 검색 시스템을 구현하기 위해, 객체 지향 분석 모델을 구성하였고 각 시스템의 기능과 논리적인 정보 표현을 위한 색인 구조를 설계하였다.

기존의 검색 방법에서 목적 정보를 모르는 경우에, 주제어에 대한 검색 결과의 목록 순서는 의미를 갖지 않는다. 따라서, 입력된 주제어에 의해 생성된 1차 정보 집합인 검색된 목록 항목을 모두 확인해야만 궁극적으로 원하는 정보를 찾게 된다. 이 논문에서는 입력된 주제어에 의해 생성된 1차 정보 집합을 목록으로 나타내는 문제점을 해결하기 위해, 1차 정보 집합이 갖는 특성을 시각적인 뷰를 통해 보여줌으로써 사용자가 원하는 정보에 관련된 부분 집합을 다음 작업의 대상으로 한다. 따라서 정보 영역을 Facet 단위로 분류하고, 이의 특성을 시각화함으로써 사용자의 이해도를 향상시키는 것을 목적으로 한다. 이와 같은 연구를 수행함으로써 기대되는 효과는 다음과 같다. 첫째, 사용자들은 다양한 방법으로 검색 과정을 수행하기 때문에 검색 방법에 대한 융통성을 제공한다. 둘째, 관련된 정보의 분석 결과를 시각적으로 제공함으로써 사용자의 이해도를 향상시킨다. 셋째, 정보 분류 영역과 관계형 구조 영역으로 구성된 분류 구조를 기반으로 하기 때문에, 정보 관리의 용이성을 제공한다. 왜냐하면, 정보 영역에 있는 항목이 변경되더라도 관계형 모델은 영향을 받지 않기 때문에, 수시로 변하는 웹 정보를 관리하는데 매우 효과적이다. 반면에, 초기 질의를 위해 기존의 방식을 적용함으로써 항목이 조합된 경우에 대한

용어의 관련성 비중을 고려하는데 어려움이 있었다. 이 연구와 관련된 향후 연구로는 인위적으로 등록된 정보 특성뿐만 아니라, 웹 정보에 대한 특성 추출 방안과 동적으로 변하는 정보에 대한 색인 요소를 포함하는 방법이 요구된다.

## 참 고 문 헌

- [1] Romana Rao, Jan O. Pederson, Matri A. Hearst, etc., "Rich Interaction in the Digital Library," *Communication of the ACM*, Vol.38, No.4, April, 1995, pp.29-39.
- [2] Peter J. Nurnberg, Richard Furuta, John J. Leggett, Catherine C. Marshall, Frank M. Shipman III, "Digital libraries : Issues and Architectures," *Proceedings of the Digital Libraries 95 Conference*, <http://www.csdl.tamu.edu/DL95/papers/nuer-nberg.html>.
- [3] Christos Faloutsos and H. V. Jagadish. "On b-tree indices for skewed distributions," In 18th VLDB Conference, August, 1992, pp.363-373.
- [4] Hurse C, Reichenberger K, Rostek L, Streitz N, "Knowledge-based Editing and Visualization for Hypermedia Encyclopedias," *Communication of the ACM*, Vol.38, No.4, April, 1995, pp.49-51.
- [5] Koichi Takeda and Hiroshi Nomiyama, "Information Outlining and Site Outlining," *Proceedings of International Symposium on Research, Development and Practice in Digital Libraries*, Nov., 1997, pp.99-106.
- [6] 홍기형, 박치향, "전자 도서관의 요소 기술 : DBMS와 정보 검색", *정보과학회지*, 제15권 제2호, 1997, pp.13-19.
- [7] Rao R, Card Sk, Jellinek HD, Mackinaly JD, Robertson GG, "The Information Grid : a framework for information-retrieval centered applications," In *Proceedings of the Acm Symposium on User Interface Software and Technology*, 1992, pp.23-32.



[8] G. G. Robertson, S. K. Card and J. D. Mackinlay, "Information Visualization using 3-D Interactive Animation," Communication of the ACM, Vol.36, No.4, April, 1993, pp.57-71.

[9] Card, Stuart K., Robertson, George G., York, William, "The WebBook and the Web Forager : An Information Workspace for the World-Wide Web," CHI96 Electronic Proceedings, 1996.

[10] <http://www.trlibn.co.jp/projects/s7710/dl/io/paraE.html>.

[11] Bruce Schatz and Hsinbun Chen, "Digital Libraries : Technological Advances and Social Impacts," IEEE Computer, Feb., 1999, pp.45-50.

[12] Comp, Introduction to the CR Classification System, Prentice-Hall, 1985.

[13] Prieto-Diaz R. and Freeman P., "Classifying Software for Reusability," IEEE Software, Vol.4, No.1, Jan., 1987, pp.6-16.

[14] Ramamoorthy C., Garg V., and Prakash A., "Support for Reusability in Genesis," IEEE Trans., on Software Engineering, July, 1988.

[15] J. Rumbaugh, etc., Object-Oriented Modeling and Design, Prentice-Hall International Editions, 1991.



**김영란**

e-mail : yrkim@cccc.chch.ac.kr  
 1988년 충북대학교 전산통계학과 (이학사)  
 1989년~1991년 충북대학교 대학원 전자계산학과(이학석사)  
 1997년 충북대학교 대학원 전자계산학과(이학박사)  
 1994년~현재 충청대학 컴퓨터학부 조교수  
 관심분야 : 정보시각화, 객체지향 소프트웨어, 정보검색



**구연설**

e-mail : yskoo@cbucc.chungbuk.ac.kr  
 1964년 청주대학교 상학과 졸업  
 1975년 성균관대학교 경영대학원 전자자료처리학과 졸업 (경영학석사)  
 1981년 동국대학교대학원 통계학과 졸업(이학석사)  
 1988년 광운대학교대학원 전자계산학과 졸업(이학박사)  
 1979년~현재 충북대학교 컴퓨터과학과 교수  
 충북대학교 전자계산소장, 한국정보과학회이사, 전산교육연구회위원장, 충청지부장, 부회장 역임  
 관심분야 : 소프트웨어공학, 정보통신, 알고리즘 등