

# 정보 시각화를 위한 웹 인터페이스 설계

## Design of Web Interface for Information Feature Visualization

홍기채(G.C. Hong) 정보유통연구팀 선임연구원  
문병주(B.J. Moon) 정보체계연구팀 선임연구원  
이성용(S.Y. Lee) 정보유통연구팀 연구원

정보 특성 시각화는 방대한 인터넷 정보로부터 사용자가 원하는 정보에 대한 이해도를 높임으로써 정보 검색시 요구되는 시간과 노력을 감소시키며, 검색된 결과에 대한 통계치를 시각적으로 보여줌으로써 검색 과정중에 효과적으로 필요정보를 탐색할 수 있게 한다. 최근 정보 시각화에 관한 관심이 증가하면서 주로 선진국을 중심으로 여러 가지 시각화 기법에 대한 연구와 이를 적용한 검색 시스템들이 개발되고 있다. 따라서 본 고에서는 정보의 특성을 체계적이면서 시각적으로 표현하기 위해, 정보 시각화의 구성 요소를 획일화하고 적합한 검색 기법을 제안하여 새로운 차원의 패러다임을 제시하고, 사용자가 보다 쉽게 정보를 향해하고 정보에 대한 이해도를 향상시키기 위한 검색 도구를 개발함으로써 검색 정보에 대한 사용자 이해도의 향상 및 효율적인 검색을 하기 위한 것이다.

## 1. 서론

최근 인터넷 사용자의 급격한 증가와 정보의 대량화 및 다양화로 인하여 사용자가 원하는 정보를 검색하기 위한 다양한 검색 기법과 검색 엔진들이 개발되고 있다. 특히 대규모적이면서 다양한 정보를 제공하는 웹 검색의 경우, 사용자들은 효율적이면서 용이하게 정보를 찾고 활용할 수 있는 강력하고도 새로운 도구를 요구한다. 대부분의 인터넷 사용자는 검색 엔진의 기능을 배워 질의어를 만들기보다는 원하는 정보를 시각화된 인터페이스를 통해 보다 편리하게 찾기를 원한다.

Huser에 의하면, 검색 엔진을 개발하는 주목적 중의 하나는 사용자에게 융통성 있고 사용자 중심으로 정보를 액세스(access)하여 사용할 수 있도록 기회를 제공하는 것이다[1]. 이를 위해 각 응용분야에 적합하게 정보 객체들을 분류하고, 분류된 객체들

을 효과적으로 화면에 시각화하여야 한다[2]. 컴퓨터 화면은 작고 2차원적이기 때문에 총 분류를 한 눈에 파악하기 어렵다. 따라서 여러 가지 분류법을 적용하여 한번에 다량의 정보를 표현할 수 있도록 한다[3-4].

대량의 웹 정보로부터 사용자가 원하는 정보를 효율적으로 검색하도록 지원하기 위해서 요구되는 연구로는 정보를 분석하고 특성을 요약하여 표현하는 기법, 정보 접근과 시각화 연구, 정보 서비스를 지원하기 위한 미들웨어 연구 등이 필요하다[5-6].

현재 Xerox PARC에서는 사용자의 정보에 대한 이해도를 향상시키기 위해 3차원 기법과 애니메이션 기법을 적용한 시각화 인터페이스를 개발하였고, 일본의 IBM 연구소에서는 정보 검색의 원리와 브라우징 기법을 통합한 정보 아웃라이닝(Outlining)을 이용하여 정보 향해 시스템(Information Navigation System: INS)을 개발하였으나 국내에서는 아직 미흡한

실정이다.

본 고에서는 ETRI에서 개발중인 검색 시스템의 통합 데이터베이스를 대상으로 하여, 정보의 특성을 체계적이면서 시각적으로 표현하기 위해 정보 시각화의 구성요소를 획일화하고 새로운 검색 기법을 적용하기 위한 패러다임을 제시하여, 사용자가 보다 쉽게 정보를 항해하고 정보에 대한 이해도를 향상시킬 수 있는 도구를 개발한다.

본 고의 구성은 다음과 같다. 먼저 II장에서는 정보 시각화 기법을 적용한 예를 제시하고, III장에서는 분야별 관계를 정의하고 표현하기 위해 정보가 갖는 특성의 윤곽을 효율적으로 시각화하는 정보 모델을 구성하고, 다중 뷰(view)와 뷰간의 동기화를 지원하는 인덱스 메커니즘을 제안한다. IV장에서는 UML(Unified Modeling Languages)을 이용하여 제안된 이론적인 개념을 적용한 정보 특성 시각화 시스템(Information Feature Visualization System)의 Use Case 다이어그램, System Sequence 다이어그램, 그리고 Design Class 다이어그램을 구성하고, V장에서는 시스템의 프로토타입을 구축한다. 마지막으로 VI장에서 결론을 맺는다.

## II. 정보 시각화 기법의 적용

### 1. Butterfly view

Butterfly view는 서비스에 대한 비동기적인 요청(request)을 제어하고 3차원 화면을 적용하기 때문에 도서 목록 리스트를 매우 효과적으로 제시해준다[7-8]. 여기에서 사용되는 GAIA 프로토콜은 메타 정보와 메타 정보 내의 다양한 자원에 쉽게 액세스하기 위한 인터페이스를 정의한 것이다. 데이터베이스에 대한 질의는 많은 시간을 필요로 하기 때문에 요청을 그래픽 객체로 구체화한다.

사용자는 질의를 통해 관심있는 주제 영역에 있는 정보를 찾을 수 있고, 질의 결과는 피라미드 형태로 표현된다. GAIA 프로토콜은 결과를 나열하는 데 있어서 제어의 역할을 하며, Butterfly는 점차적으로

증가하는 결과를 나타내준다. 이러한 정보는 각기 다른 데이터베이스에 저장된 정보로부터 검색된 결과인데 반해서, 독립적으로 존재하지는 않는다. 따라서, 사용자는 질의를 통해 속성을 갖는 정보들간의 관계로부터 결과를 얻는다. Butterfly view는 정보들간의 관련 상태를 시각화하여 이들간의 상호 작용을 통해 질의를 할 수 있는 환경이다.

### 2. Scatter/Gather

Scatter/Gather 브라우징은 주제목과 대표적인 항목을 기초로 하여, 사용자가 특별한 관심이 있는 데이터 항목의 부분 집합을 찾도록 허용한다[9]. 즉, 대량의 정보 집합을 항해하는 사용자에게 도움을 주기 위해, 동적인 table-of contents를 사용한다. 초기의 질의에 대한 정보 클러스터링을 검색한 후, 클러스터링 집합을 작은 그룹으로 자동적으로 Scatter한다. 이 Scatter된 그룹들에 대한 간략한 요약물 사용자에게 보여준다.

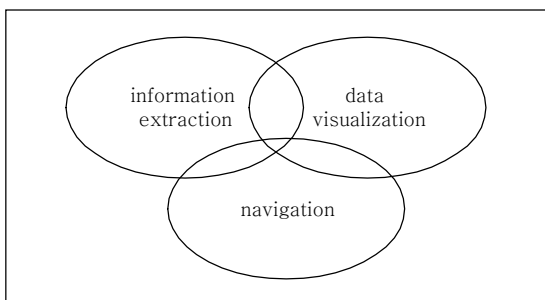
사용자는 요약 항목 중에서 하나 또는 여러 개의 그룹을 선택한다. 선택된 그룹은 합해지고, 이러한 반복적인 작업에 의해 그룹은 더욱 더 작아지고 상세화된다. 적용된 정보 클러스터링 알고리즘은 linear-time 클러스터링 알고리즘으로서, 정확성보다는 반복을 활성화하기 위해 속도를 최적화하는 것을 목적으로 하고, 자료 구조는 계층 구조를 적용한다.

### 3. Information Outlining

디지털 도서관은 방대한 양의 문자 자료와 멀티미디어 자료 때문에 정보 요구의 범위가 넓고 많은 것들이 term-based 질의로 표현될 수 없다. 예를 들면, 수십만 개의 신문 기사 중에서 중요한 정보를 검색하려면 straightforward scanning이나 term-based searching(키워드와 전문 검색)으로는 충분치 않다. Straightforward scanning은 너무나 많은 시간이 소모되고 term-based searching은 기사의 정확한 위치를 찾는 데 제한적이다.

디지털 도서관에 대한 정보 아웃라이닝은 다중

차원화 관점에서 다른 종류의 대규모 데이터 집합을 검색하고 접근하도록 하는 새로운 기법이다[6, 10]. 정보 아웃라이닝은 (그림 1)과 같이 자료를 통한 정보 추출(information extraction), 자료 시각화(data visualization), 그리고 항해(navigation)를 통합한 기법으로서, 데이터의 특정 부분을 가시화하기 위해 다중 뷰를 제공한다.



(그림 1) 아웃라이닝의 개념

전형적인 정보 검색 기법과 정보 아웃라이닝 기법을 비교하면 <표 1>과 같다.

<표 1> 정성적인 성능 분석

	이미 알려진 문서 찾기	모르는 문서 찾기	주제에 대한 정보 수집	관심있는 정보발견
전형적인 검색	***	**	**	*
정보 아웃라이닝	***	***	****	****

\*: Poor, \*\*: Ordinary, \*\*\*: Good, \*\*\*\*: Excellent

전형적인 검색 방법에서 목적 정보를 모르는 경우에, 주제어를 입력하여 매칭되는 정보에 대한 검색결과와 목록은 정확한 의미를 갖지 못한다. 따라서 입력된 주제어에 의해 생성된 1차 정보 집합인 검색 목록 항목을 모두 확인해야 한다. 그러나 정보 아웃라이닝 기법은 입력된 주제어에 의해 생성된 1차 정보 집합을 단순히 목록으로만 나타내기보다는, 1차 정보 집합이 갖는 특성을 시각화한 뷰를 통해 보여줌으로써 사용자가 원하는 정보에 관련된 부분 집합을 선택할 수 있다. 따라서 작업 영역을 단계적

으로 부분집합화 하고 이의 특성을 시각화함으로써 사용자의 이해도를 향상시킬 수 있다.

#### 4. 비교 분석

정보 아웃라이닝 기법은 정보 시각화와 항해에 중점을 둔 브라우징 시스템으로 관계형 인덱스 구조만을 사용하기 때문에 정보간의 다양한 관련성을 정의하는 데 제한점이 있다.

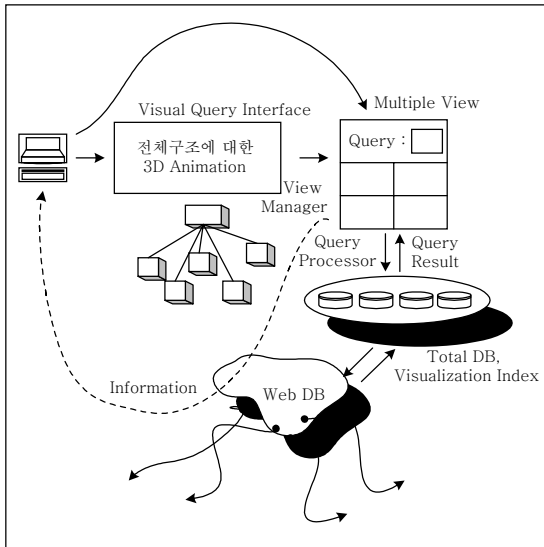
반면에 Scatter/Gather 방법에서의 클러스터링 기법은 사용자가 원하는 정보의 부분 집합을 효율적으로 찾으려 지원하기 때문에 대량의 정보를 검색하는 경우에 매우 효과적일 수 있으나 클러스터링을 하기 위한 정보의 시맨틱스를 구성하는 것이 문제이다.

### III. 시각화 지원 웹 검색 시스템

시각화 지원 웹 검색 시스템은 윈도우 프로그램 간의 호환성을 유지하는 일관성 있는 디자인, 사용자의 이해도를 높이기 위해 복잡성을 배제한 간략한 화면 구성, 정보의 의미를 명확히 하기 위한 명료성에 중점을 둔 시각화 인터페이스로서, 여기에서는 정보 시각화 웹 검색 시스템의 전체적인 구조를 기술하고 관심있는 정보의 식별과 분류 방법을 기술한다. 또한 분류된 정보를 검색하기 위한 인덱스 메커니즘을 제시한다.

#### 1. 시스템 개요

정보 특성 시각화 시스템은 정보 아웃라이닝 기법을 적용한 시각화 웹 검색 시스템으로 (그림 2)에 서와 같이 정보 검색을 위한 효율적인 정보를 다중 데이터 소스(multiple data source)로부터 속성 값 쌍(attribute-value pairs)을 추출한다. 속성 값의 쌍인 데이터는 각 뷰에 대한 모델에 의해 변환되고, 각 뷰를 통해 상호 작용이 가능하도록 시각화된다. 뷰 계층에서는 데이터 집합에 대한 아웃라이닝을 뷰에서 사용자에게 보여주고 사용자는 뷰를 통해 상호 작용한다.



(그림 2) 시스템의 구성

## 2. 정보 식별 및 분류

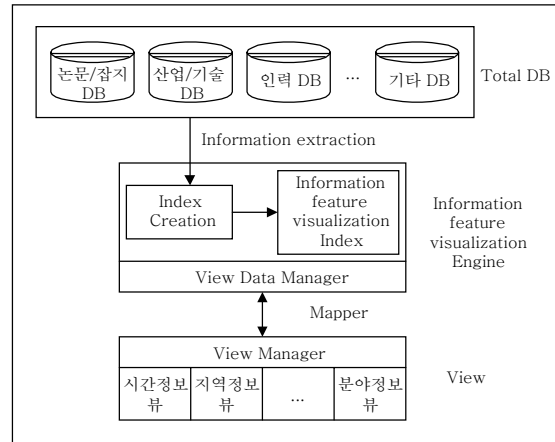
전형적인 정보 검색 기법이 데이터베이스로부터 사용자가 원하는 정보를 찾는 데 목적을 둔 사용자 중심이라면, 정보 특성 시각화의 목적은 데이터베이스에 무엇이 있는지를 보여주기 위한 것으로 자료 중심의 정보 검색 기법이다.

정보 검색(retrieval), 정보 여과(filtering), 하이퍼텍스트와 같은 정보 획득(acquisition)에는 몇 가지 패러다임이 있다.

(그림 3)은 정보 아웃라이닝 패러다임을 가능하게 하는 모델이다. 정보 특성 시각화 계층은 정보 검색을 위한 효율적인 정보를 multiple data source로부터 속성 값 쌍으로 추출한다. 속성 값의 쌍인 데이터는 각 뷰에 대한 모델에 의해 변환되고, 각 뷰를 통한 상호 작용을 가능하게 하고 시각화한다. 뷰 계층에서는 데이터 집합에 대한 아웃라이닝을 뷰에서 사용자에게 보여주고 사용자는 뷰를 통해 상호 작용한다.

## 3. 정보 모델

웹 정보를 대상으로 한 인터페이스 개발은 정보의 특성이 반영된 정보 공간에 대한 구조적 정보 모



(그림 3) 정보 아웃라이닝 모델

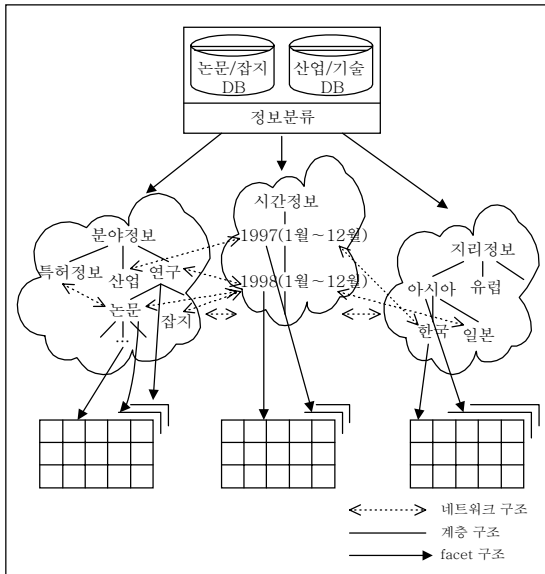
델을 정립할 필요가 있다.

여기에서는 정보간의 관계성을 가장 잘 표현하는 관계형 모델을 기반으로 하여, 정보의 공통 특성을 반영하도록 소프트웨어 재사용 기법인 facet 기법과 구조적 관계 분류 기법을 적용한다. Facet 기법은 정보의 특성에 대한 계층적 표현이 곤란한 경우에 적용하는 기법으로서, 관련성에 의한 분류가 아니라 정보의 공통적인 특성에 따른 표현 모델이다[11]. 일반적으로 facet 기법은 <표 2>와 같이 정보의 특성을 기능적 facet과 환경적 facet으로 분류한다.

<표 2> 정보의 특성에 대한 facet 분류

기능적 facet	환경적 facet
분야정보	시간정보 지리정보

Facet 기법의 기본 원리에 따라 정보의 공통적인 특성을 고려하여 정보를 분류한 후, 관계형 모델을 기반으로 한 구조적 관계 분류 방법에 의해 정보 그룹간의 관련성을 정의하고자 기존의 데이터베이스에 저장된 자료들을 (그림 4)와 같이 분류한다. 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 정의 테이블이다[11-12].



(그림 4) 정보 분류 구조

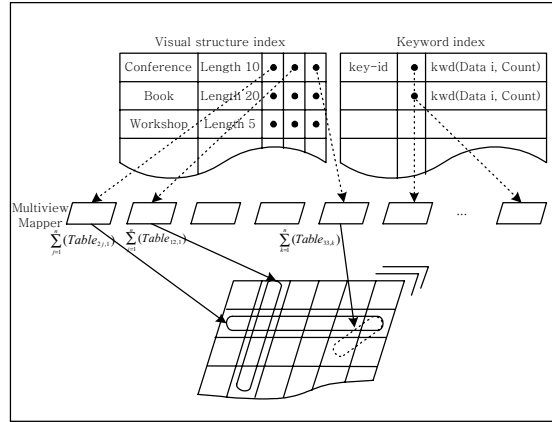
논문/잡지 데이터베이스와 산업/기술 데이터베이스의 공통 필드를 기반으로 하여 분야정보 facet, 시간정보 facet, 지리정보 facet으로 분류한 후, 각 facet에 포함된 정보들의 논리적인 표현으로 관계형 테이블 구조를 적용하여 표현한다.

또한, 뷰간의 이동을 가능하도록 하기 위해 facet 간의 관련성을 네트워크 구조와 계층 구조로 표현한다. 이러한 정보간의 관련성은 포인터를 통하여 네트워크 구조에 연결시키거나, 또는 3~4개의 계층 구조에 연결시킨다. 연결된 정보에 대해서는 다중 뷰를 구성하도록 하기 위해 관련성을 정의하고, 이에 대한 테이블과 SQL 코드를 작성한다.

#### 4. 인덱스 메커니즘

데이터의 정형적인 특성에 대한 통합 데이터베이스의 인덱스 구조를 바탕으로 하여 정보 시각화를 지원하기 위해서는 데이터가 갖는 비정형적인 특성과 관계성에 대한 인덱스 구조가 요구된다. 이를 위해 (그림 5)와 같은 인덱스 메커니즘 구조를 제시한다.

기존에 정의된 통합 인덱스는 Keyword index로서, key-id에 대한 데이터베이스의 속성을 구성 요소로 한다. Visual structure index는 데이터의 비정형



(그림 5) 인덱스 메커니즘

적인 필드로부터 데이터의 시각화 관점을 위한 속성을 추출하여 구성한 것으로서, 뷰를 통한 사용자 질의에 대해 뷰 관리자와 매핑 과정(mapping operation)이 작동하고 저장소에 저장된 데이터간에 데이터링크 과정(data linking operation)이 작동한다.

다중 뷰에 대한 관련성을 갖는 Visual structure index를 구성함으로써 정보가 갖는 여러 가지 특성을 다중 뷰를 통해 동기적으로 시각화 할 수 있다. 따라서, 검색 환경에 대한 사용자의 이해도를 확장시켜 검색의 효율성을 향상시킨다.

정보 아웃라이닝을 위한 필수적인 기능은 키워드 추출, 키워드 정형화, 키워드 목록화를 통해 원시 데이터로부터 특징을 추출하고, 통계치(statistics)와 군집화(clustering)를 통해 데이터 집합에 대한 속성 값 쌍으로 변환하고, 결과를 시각화하며 subset selection zooming, media scaling, media conversion을 통해 상호 작용이 가능하도록 한다.

#### 가. 링커

이와 같은 기능을 수행하기 위한 메커니즘으로 뷰의 물리적인 표현을 관리하는 뷰 관리자와 인덱스 테이블에 있는 데이터간의 매핑을 위한 링커(linker)가 필요하다.

뷰 관리자는 시각적 표현을 사용하여 정보의 특성을 강조하는 것으로서 생성, 폐쇄, 검색 단어 설정,

읽기, 정보 획득 등의 연산 기능을 지원한다. 또한 뷰 간의 관련성을 추적하여 뷰간의 동기화를 가능하도록 매핑시키는 기능을 수행한다.

나. 매핑

뷰와 시각화 구조 인덱스 테이블간의 매핑과 정보에 대한 관계형 테이블간의 매핑을 계산하는 테이블은 <표 3>과 같다.

<표 3> 매핑 테이블

뷰	뷰 객체	속성
시간정보 뷰	Bar(height) Bar(position)	Count(articles) Date <sub>i</sub>
지역정보 뷰	Button(label) Button(label)	Count(articles) Location <sub>i</sub>
분야정보 뷰	Button(label) Button(color)	Count(articles) Category <sub>i</sub>

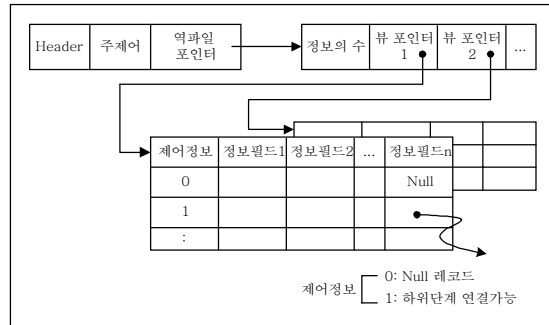
사용자는 뷰를 이용하여 질의를 할 수 있으며, 정보 특성 시각화 시스템은 뷰의 그래프를 이용하여 질의 결과를 시각적으로 표현할 수 있다. 이때, 질의 결과는 뷰에 따라 각기 다른 형태의 그래프로 나타나며, 사용자는 표현된 그래프 중에서 원하는 부분을 선택함으로써 제한된 질의를 추가한다.

예를 들어, 시간정보 뷰의 경우 막대 그래프의 위치(position)는 시간(년/월)을 나타내며, 높이(height)는 날짜(Date)란 속성으로부터 구한 데이터의 합계를 의미한다. 마찬가지로 지역정보 뷰의 경우엔 위치(Location<sub>i</sub>) 속성에 따라 표현된 지역의 정보를 나타내며, 분야정보 뷰의 경우엔 분야(Category<sub>i</sub>) 속성에 의한 정보의 크기를 나타낸다.

다. 인덱스 파일

인덱스 파일은 저장된 정보의 주제어와 비정형적 특성을 나타내는 레코드 및 레코드들과 관련된 정형적인 정보의 위치정보인 포인터를 저장한다. (그림 6)은 정보 특성 시각화 브라우저를 통해 자료 검색이 가능하도록 한 인덱스 파일 구조이다.

뷰 포인터에 연결된 테이블은 하위 작업 여부를



(그림 6) 뷰 검색을 위한 인덱스 파일 구조

표현하는 제어 정보 필드를 갖는다. 연도에 따른 월 별 정보가 저장되어 있는 경우는 제어 정보 필드의 값이 1이 된다.

IV. UML을 이용한 시스템 모델링

여기에서는 UML을 이용하여 정보 시각화 웹 검색 시스템에 대한 모델링을 한다[13-14]. 먼저 웹 검색 시스템의 작동에 대한 시나리오를 기반으로 Use Case 다이어그램을 작성하고, 이를 이용하여 System Sequence 다이어그램과 Class 다이어그램을 작성한다.

1. 시나리오

이 절에서는 시스템의 작동 과정을 서술식으로 기술한 시나리오를 작성한다. 이것은 앞으로 작성할 여러 다이어그램에 대한 기초자료로서 이 이외에도 CRC-card나 비기능적 요구사항(Non-Functional Requirement)을 추가하여 기술할 수 있다.

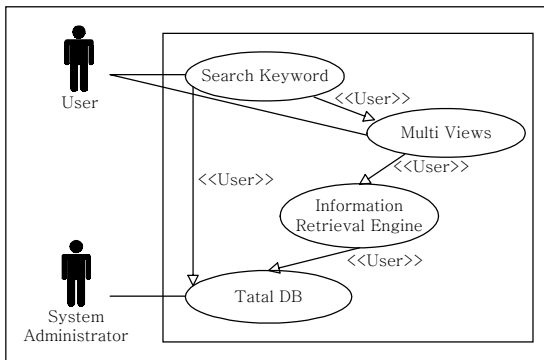
다음은 정보 시각화 웹 검색 시스템에 대한 시나리오이다.

정보 시각화 웹 검색 시스템의 사용자는 먼저 관심있는 자료에 대한 키워드를 입력한다. 다중 뷰에서는 검색된 자료의 통계치를 시각적으로 각 자료의 특성에 해당되는 뷰를 통해 표현한다. 이때 뷰 관리자는 뷰간의 동기화와 항해를 관리하며, 질의 관리자는 각 뷰로부터의 질의 연산을 수행하여 검색하는

부분으로 넘겨주고, 검색된 질의 결과를 각 뷰로 전달해주는 역할을 수행한다. 각 뷰는 검색된 자료를 표현하는 역할뿐 아니라 검색된 자료에 대한 부분 검색을 수행한다. 이것은 각 뷰에 속해있는 버튼과 같은 컴포넌트에 의해 수행된다.

## 2. Use Case 다이어그램

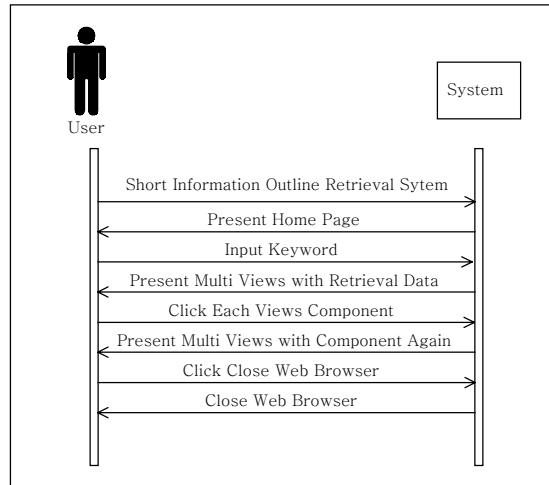
정보 시각화 웹 검색 시스템에 대한 시나리오를 기반으로 UML로 작성한 Use Case 다이어그램은 (그림 7)과 같다. 이 시스템에서 사용되는 Use Case는 크게 4부분으로 구성된다. Search Keyword Use Case는 사용자로부터 입력받은 키워드를 검색하는 부분이고, 검색된 결과는 Multi Views Use Case의 정보 특성에 따라 시각적으로 표현된다. 이때 사용자는 Multi Views를 통해 재질의 할 수 있다. Total DB Use Case는 웹상에서 검색된 자료에 대한 정보를 저장해 놓은 통합 DB로서 시스템 관리자에 의해 관리되며, Information Retrieval Engine Use Case는 키워드 검색이나 다중 뷰를 통해 전달된 질의어를 이용해 통합 DB로부터 요구한 자료를 검색하는 부분이다.



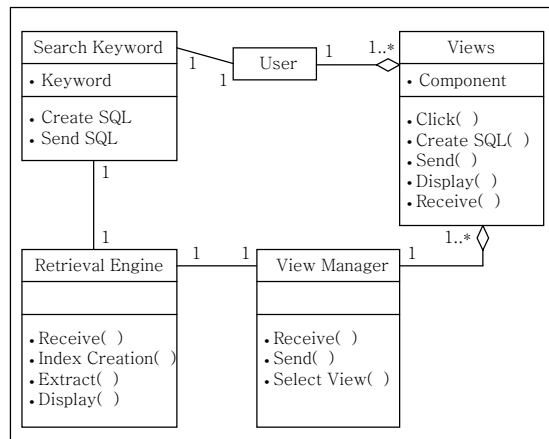
(그림 7) Use Case 다이어그램

## 3. System Sequence 다이어그램

(그림 8)은 사용자와 시스템간의 상호작용을 나타낸 Sequence 다이어그램으로 시스템의 동적 측면을 표현한다.



(그림 8) System Sequence 다이어그램



(그림 9) Class 다이어그램

## 4. Class 다이어그램

시나리오를 기반으로 추출한 클래스들을 표현한 Class 다이어그램은 (그림 9)와 같다. 사용자는 키워드 검색 클래스와 1대 1 관계를 가지며 키워드 검색 클래스는 속성으로 키워드를 갖고, 함수로는 키워드를 이용해 SQL 질의문을 생성해 검색엔진으로 보낸다. 또한 사용자는 한 개 이상의 뷰를 통해 질의 및 질의 결과를 확인할 수 있으며, View를 통한 질의는 질의 관리자에 의해 관리된다. 질의 관리자는 View로부터 요청된 질의를 검색 엔진으로 보내며, 검색 결과를 다시 View로 넘겨주는 중개 역할을 한다.

## V. 구현

이 시스템은 웹 환경에서 동작하며, C 언어를 사용하여 구현하였고, 시스템의 데이터 오브젝트는 통합정보검색시스템에서 사용하는 통합 데이터베이스 스키마를 참조하여 작성되었다. 인포믹스 DB로 작성된 테이블로부터 자료를 검색하기 위한 질의 처리는 fulcrum의 SearchSQL을 이용하여 키워드 검색과 뷰를 이용한 검색이 이루어진다.

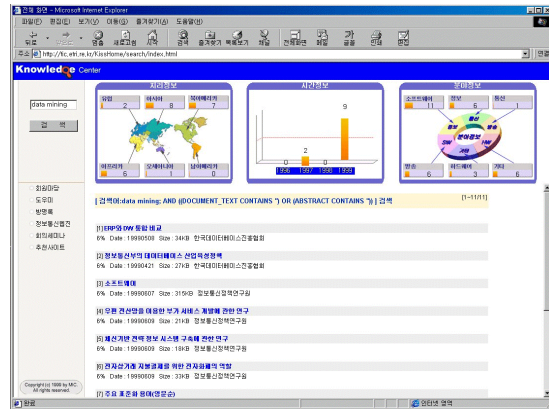
이 시스템의 사용자는 키워드 검색과 뷰를 이용한 검색 등 다양한 방법으로 질의를 구성할 수 있으며, 이때 선택된 질의 형태에 따라 새로운 테이블이 형성된다. 사용자가 질의한 결과를 출력하면서 지리 정보 뷰, 시간정보 뷰, 분야정보 뷰의 내용도 함께 출력하게 된다. 뷰를 이용한 추가 질의가 주어지면 검색된 정보 내에서 새로운 질의 결과와 각각의 뷰들이 재구성되어 출력된다.

이들 뷰는 데이터의 특정 속성을 시각적으로 표현한 것으로, 각 뷰들은 서로 동기화가 이루어진다. 또한 이들 뷰들을 그래픽적으로 표현함으로써 초보자들도 손쉽게 원하는 정보에 접근할 수 있게 하였다. 여러 개의 뷰들 중에서 일반적으로 유용하게 이용될 수 있는 지리, 시간, 분야정보만을 우선적으로 뷰로 구성하였다.

(그림 10)은 정보 시각화 웹 검색 시스템의 프로토타입에 대한 초기화면이다.

## VI. 결론

본 고에서는 정보 검색 기법과 브라우징 기법을 통합한 웹 정보 시각화 인터페이스 시스템을 설계하고 이를 바탕으로 프로토타입을 구축하였다. 이를 위해, 정보가 갖는 특성을 효율적으로 시각화하기 위한 구성 요소간의 관계를 정립한 정보 모델을 구성하고, 정형적인 정보와 비정형적인 정보에 대한 동기화를 지원하는 인덱스 메커니즘을 제안하였다. 또한, UML을 이용하여 정보 시각화 웹 검색 시스템에 대한 Use Case 다이어그램, System Sequence



(그림 10) 정보 시각화 웹 검색 시스템

다이어그램, 그리고 클래스 다이어그램을 구성한 후 프로토타입을 구축하였다. 인덱스 메커니즘은 문헌 색인의 형태인 역과일 구조와 여과의 개념을 기반으로 한 요약 파일 구조를 이용하였다.

기존의 웹 정보 검색 기법이 데이터베이스로부터 저장된 정보를 찾는 데 중점을 둔 반면에, 본 고에서는 데이터베이스에 저장된 정보가 무엇이 있는지를 효율적으로 시각화 함으로써 사용자의 관점에 중점을 두었다. 정보의 특성을 효율적으로 시각화하는 시스템을 구성하기 위한 정보 모델 구성과 인덱스 메커니즘에 관한 연구를 수행함으로써 기대되는 효과는 다음과 같다. 첫째, 방대한 인터넷 정보에 대한 효율적인 시각화 관점에서 볼 때, 사용자가 원하는 정보에 대한 이해도를 높임으로써 정보를 찾는 데 요구되는 시간과 노력을 감소시킬 수 있다. 둘째, 웹 환경을 이용하는 전자 도서관 시스템과의 통합 관점에서 볼 때, 도서 정보 및 통계 자료에 대한 시각화를 지원함으로써 검색 과정에서 정보에 관련된 동향을 파악할 수 있다.

지금까지의 연구는 웹상에서 검색된 문서 자료를 그 특성에 따라 2차원으로 시각화하여 보여주고 있다. 향후 연구로는 기존의 정보 의미를 기반으로 한 정보 추출 및 클러스터링 기법에 대한 연구와 개방형 분산 웹 환경에서 요구되는 웹 엔진 구조에 대한 연구 및 3차원을 통한 정보 시각화와 멀티미디어 검색이 추가되어야 한다.



## 참고 문헌

- [1] C. Huser, K. Reichenberger and L. Rostek, N. Streitz, "Knowledge-based Editing and Visualization for Hypermedia Encyclopedias," *Communications of the ACM*, Vol. 38, No. 4, Apr. 1995, pp. 49 – 51.
- [2] Fox EA, Hix D, Nowell LT, Brueni DJ, Wake WC, Health LS, Rao D. Users, "*User Interface and Objects: Envision, a Digital Library*," *Journal of the American Society for Information Science*, Vol. 44, No. 8, Sep. 1993, pp. 480 – 491.
- [3] Peter J. Nurnberg, Richard Furuta, John J. Leggett and Catherine C. Marshall, Frank M. Shipman III, "Digital Libraries: Issues and Architectures," *Proc. of the Digital Libraries 95 Conference*, <http://www.csdl.tamu.edu/DL95/papers/nuernberg.html>.
- [4] Romana Rao, Jan O. Pederson and Marti A. Hearst, etc. "Rich Interaction in the Digital Library," *Communications of the ACM*, Vol. 38, No. 4, Apr. 1995, pp. 29 – 39.
- [5] D. Lucarella and A. Zanzi, "A Visual Retrieval Environment for Hypermedia Information System," *ACM Trans. on Information Systems*, Vol. 14, No. 1, Jan. 1996, pp. 3 – 29.
- [6] Koichi Takeda and Hiroshi Nomiyama, "*Information Outlining for Digital Libraries*," Research Report-RT0181, IBM Research, Tokyo Research Laboratory, Jan. 1997.
- [7] 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, Apr. 1993, pp. 57 – 71.
- [8] J. Mackinlay, R. Rao and Card SK, "An Organic User Interface for Searching Citation Links," In *Proc. of ACM Conference on Human Factors in Computer System*, Addison Wesley, May 1995, pp. 67 – 73.
- [9] D.R. Cutting, D.R. Karger and J.O. Pedersen, "Constant Interaction-time Scatter/Gather Browsing of Very Large Document Collections," In *Proc. of SIGIR'93*, Pittsburgh, PA., July 1993, pp. 126 – 134.
- [10] <http://www.trl.ibm.co.jp/projects/s7710/dl/intlp.html>
- [11] R. Prieto-Diaz and P. Freeman, "Classifying Software for Reusability," *IEEE Software*, Vol. 4, No. 1, Jan 1987, pp. 6 – 16.
- [12] T.J. Lee, "Software Reuse in Parallel Programming Environments," Ph.D. dissertation, Department of Computer Science, The University of Texas at Austin, 1989.
- [13] G. Booch, J. Rumbaugh and I. Jacobson, "The Unified Modeling Language User Guide," Addison Wesley Longman, Inc., Feb. 1999.
- [14] Craig Larman, *Applying UML and Patterns*, Prentice Hall, 1998.