

정보 특성 시각화 시스템 구축

(Implementation of an Information Feature Visualization System)

조 윤 기 [†] 하 재 관 ^{**} 구 연 설 ^{***}
 (YoonKi-Cho) (JaeKwan-Ha) (YeonSeol-Koo)

요 약 정보 특성 시각화는 기존의 정보 검색 기법과 시각화 기법을 통합하여 방대하고 다양한 인터넷 정보에 대한 이해도를 높임으로써 정보 검색 시 요구되는 시간과 노력을 감소시키며, 검색된 결과에 대한 통계치를 시각적으로 보여줌으로써 검색 과정 중에 정보에 관련된 동향을 파악하는데 유용하다. 전자 도서관의 예에서 볼 수 있듯이 최근 정보 시각화에 관한 관심이 증가하면서 주로 선진국을 중심으로 여러 가지 시각화 기법에 대한 연구와 이를 적용한 검색 시스템들이 개발되고 있다. 따라서 이 논문에서는 정보의 특성을 체계적이면서 시각적으로 표현하기 위해, 정보 시각화의 구성 요소를 다양한 뷰를 통해 확일화하고 구조적 정보 분류 기법을 적용한 새로운 차원의 패러다임을 제시하고, 사용자가 보다 쉽게 정보를 향해하고 정보에 대한 이해도를 향상시키기 위한 검색 도구를 개발함으로써 검색 정보에 대한 사용자 이해도와 검색의 효율을 향상시켰다.

Abstract Information Feature Visualization integrates the traditional information retrieval techniques and visualization techniques and reduces the time and effort from searching for requested information by promoting apprehension from the wide and various internet information. It's used to understand the related information tendency by supporting the statistic materials for retrieved results. It's useful to catch a tendency regarding information over the search process. Recently, by increasing interest in the information visualization in the library system, The research about various visualization technology and searching system has been developed by many researchers. So, to describe systematically and visually the feature of information, we'll make the element of information visualization uniform by various views and suggest paradigm renewing retrieval technology using information outlining techniques by developing search tool to improve apprehension about information and navigate more easily among information. We can improve user's apprehension about search information and search more usefully.

1. 서 론

정보 통신의 발달과 더불어 인터넷 사용이 급격히 증가함에 따라 사용자가 원하는 정보를 효율적으로 찾고 활용할 수 있는 검색 도구가 필요하게 되었다. 대부분의 검색 엔진을 통한 검색은 텍스트 형태의 질의 및 결과

표현을 이용하므로 이를 이용하는 사용자들은 검색 엔진의 기능 및 사용 방법을 익히고 많은 검색 결과 목록으로부터 관심 정보를 찾기보다는 원하는 정보를 시각화된 인터페이스를 통해 보다 편리하게 찾기를 원한다 [1][2]. 전자 도서관의 예에서 볼 수 있듯이 최근 정보 시각화에 대한 관심이 점점 높아지고 있다[3][4].

기존의 정보 검색 기법을 이용한 검색 환경은 정보 처리 및 효율성 면에서 다음과 같은 문제점이 있다. 첫째, 대용량의 데이터를 대상으로 검색이 이루어지므로 너무 많거나 부정확한 검색 결과를 얻는다. 따라서 사용자가 원하는 정보를 얻기 위해서는 또 다른 검색을 하지 않으면 안된다. 둘째는 사용자 자신의 요구와 일치하는 정확한 질의를 표현하기가 쉽지 않다. 이러한 문제를

[†] 학생회원 : 충북대학교 전자계산학과

ykcho@selab.chungbuk.ac.kr

^{**} 비 회 원 : 우송공업대학 무역사무자동화과 교수

jgha@selab.chungbuk.ac.kr

^{***} 종신회원 : 충북대학교 컴퓨터학과 교수

yskoo@cbucc.chungbuk.ac.kr

논문접수 : 2000년 1월 5일

심사완료 : 2000년 7월 10일

해결하기 위해서 사용자는 데이터 집합의 특징과 정보 검색이 제공하는 질의의 형태에 대한 정보를 알고 있어야 한다. 마지막으로 검색 대상이 되는 데이터의 전체 구조와 상호 관련성을 효과적으로 제시하지 못한다.

이러한 문제를 해결하기 위해 이 논문에서는 정보의 공통 특성과 이들간의 관련성을 효율적으로 시각화하는 것을 목적으로 한다. 연구 방법으로는 ETRI의 통합 데이터베이스를 대상으로 하여, 정보의 특성을 체계적이면서 시각적으로 표현하기 위해 정보의 공통 특성과 관련성을 체계적으로 표현할 수 있는 구조적 정보 분류 기법과 색인 기법을 제안한다. 또한 제안된 기법과 브라우징 기법을 통합하여 사용자가 보다 쉽게 정보를 향해하고 정보에 대한 이해도를 향상시킬 수 있는 정보 특성 시각화 시스템을 설계 및 구현한다.

이 논문의 구성은 다음과 같다. 먼저 2장에서는 정보 시각화 시스템의 연구 동향 및 구성 요소를 기반으로 한 분류 기법을 기술하고, 3장에서는 정보의 특성을 시각화하기 위한 구조적 정보 분류 방안 및 색인 구조를 제안한다. 4장에서는 시스템의 구조 및 구성 요소의 기능을 기술하고, 5장에서는 UML을 이용하여 정보 특성 시각화 시스템(Information Feature Visualization System)을 설계한다. 6장에서는 구현 결과를 기술하고, 마지막으로 7장에서 결론을 맺는다.

2. 관련연구

정보 시각화에 대한 관심이 증가되면서 전자 도서관(digital library) 시스템을 중심으로 다양한 정보 시각화 프로젝트(Information Visualizer Project)들이 수행되고 있다. Hearst에 의한 TileBars 시스템은 문서 내의 키워드 분포 정보를 시각화 한 것으로 사용자에게 검색된 문서의 상대적인 길이, 질의에 포함된 키워드의 상대적인 빈도수와 문서 사이의 분포 정보 등을 시각적으로 제공한다[5]. 또한 Xerox PARK에서는 Butterfly View, Inforgrid 등과 같은 정보 시각화 시스템들을 개발하였다[6][7]. Butterfly View 시스템은 서비스에 대한 비동기적인 요청(request)을 제어하고 3차원 화면을 통해 도서 목록을 효과적으로 제시하며, Inforgrid 시스템은 문서 영역에 대한 시각화와 관리 측면을 강조하기 위해 영역에 대한 접근 용이성(ready-to-hand) 개념을 이용함으로써 시각화 도구로서의 유용성을 제공한다. 일본의 IBM 연구소에서는 전자 도서관의 시각화를 지원하기 위한 방안으로 다중 뷰 기법을 적용하여, 신문 기사 정보로부터 사용자가 원하는 정보를 시각적으로 검색 가능하도록 하였다[8]. 이와 같은 연구들은 정보의

체계적인 분류 방안보다는 정보 시각화를 위해 검색 기법과 브라우징 기법을 통합하는데 중점을 두고 있다. 따라서 사용자가 원하는 정보에 접근하기 위한 시각화된 환경을 제공하고 있는 반면에, 정보의 개요를 파악할 수 있는 공통적인 특성과 관련성을 시각화하지 못하는 단점이 있다.

이와 같은 단점을 해결하기 위해서는 정보의 특성이 반영된 정보 공간에 대한 체계적인 정보 분류 기법이 요구된다. 정보의 특성 및 관련성을 표현하기 위해 일반적으로 많이 사용되는 그래프 모델, 개체관계 모델, 의미망 모델 등의 경우엔 정보 표현의 제약, 정보 특성 반영 부족과 같은 문제점이 있다. 반면에 객체 지향 데이터 모델의 경우엔 비교적 유연하게 정보를 표현할 수 있지만 정보의 특성을 충분히 반영하지 못한다. 구성 요소를 기반으로 한 정보 분류 구조로는 계층구조[9]와 중심어를 기반으로 한 구조[10], 그리고 네트워크 구조가 있다. 계층 구조는 종속 관계와 포함관계의 원리를 기반으로 하고 나무 구조와 격자(lattice) 형태가 있다. 이 구조는 대량의 정보가 갖는 복잡한 관련성을 표현하기가 어렵다. 중심어를 기반으로 한 구조는 정보간의 관련성에 의한 분류가 아니라 공통적인 특성에 의한 분류로서 정보를 계층적으로 정확하게 분류하기 어려운 경우 이용된다. 이 구조는 정보의 수가 증가하는 경우에, 하나의 클래스에 너무 많은 정보를 그룹화 하는 단점이 있다. 네트워크 구조는 모든 정보를 포인터에 대한 네트워크를 통해 연결하므로 여러 개의 나무 구조를 연결하여 복합(compound) 구조를 쉽게 구성할 수 있다. 반면에, 정보의 위치를 정하기 위해 긴 경로를 추적해야 하므로 오버헤드가 발생할 수 있다.

이 논문에서는 중심어를 기반으로 한 구조적 정보 분류 기법을 통해 웹 정보의 특성에 대한 계층적 표현의 문제점을 해결하고 정보간의 효율적인 관련성을 시각화하기 위한 정보 특성 시각화 시스템을 개발한다.

3. 시스템 분석

3.1 구조적 정보 분류

이 논문에서는 정보간의 관련성을 표현하기 위해 관계형 모델을 기반으로 하여 정보의 공통 특성을 반영하도록 소프트웨어 재사용 기법인 패싯(facet)기법[10]과 구조적 관계 분류 기법을 적용한다. 패싯 기법은 정보의 특성에 대한 계층적 표현이 곤란한 경우에 적용하는 기법으로 여기에서는 정보가 갖는 공통적인 특성을 나타내는 기본 단위로 사용한다. 구조적 분류 방법[11]은 관계 구조, 네트워크 연결 구조, 그리고 계층 구조를 병합

한 형태로 관계 모델을 응용한 것으로 정보 시각화를 지원하기 위한 정보 구조를 그림 1과 같이 정보가 갖는 공통적인 특성을 표현하는 정보 분류 영역과 정보 그룹 간의 관련성을 정의한 관계형 구조 영역으로 구성한다.

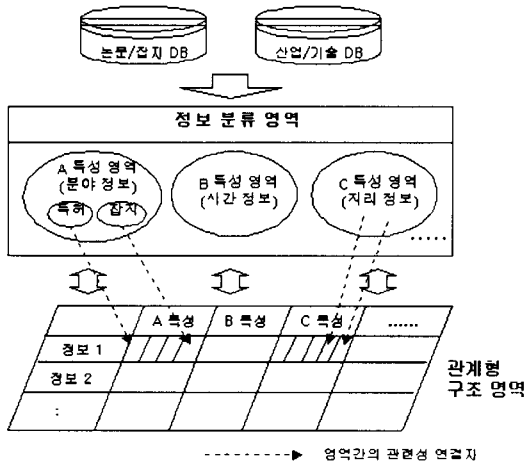


그림 1 구조적 정보 분류 방법

정보 분류 영역에서는 분야 정보를 나타내는 기능적 패킷과 시간과 지리 정보를 나타내는 환경적 패킷으로 정보 특성을 분류한다. 분류된 자료들은 분야에 따른 테이블, 시간에 따른 테이블, 지리적인 위치에 따른 테이블, 그리고 이들 테이블 간의 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 관계가 존재한다. 데이터베이스에 저장된 정보 분류 영역에 대한 집합(A특성, B특성, C특성, ...)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.

3.2 통합 데이터베이스 목록 구조

정보 특성 시각화 시스템은 Etri의 통합 데이터베이스를 대상으로 한 것으로 표 1에서 보는 바와 같이 논문/잡지, 산업/기술, 인력, R&D, 산업 통계, 국제 협력 등 다양한 데이터베이스에 대한 그룹별 통합 스키마와 공통 필드명을 구성한다. 레코드 번호는 그룹 번호(2자리) + 데이터베이스 번호(2자리) + 레코드 순서 번호(6자리)

표 1 데이터베이스별 통합 스키마와 공통필드
(가) 통합스키마

레코드번호	검색횟수	도큐먼트형태	로딩일자	색인처리	원문파일명	DB필드			제목	주제어	내용
						저자	발명자	...			
0101000001	0	HTML			/etlars93/electronic/jugidong/800					컴퓨터; 통신;	

(나) 공통필드

번호	필드명	속성명	태그번호	길이	관련필드명
1	레코드번호	recno	001	F(10)	그룹번호+DB번호+레코드일련번호
2	검색횟수	accesscnt		F(8)	
3	도큐먼트형태	docform	002	F(4)	
4	로딩일자	date		F(12)	발행일자,개최일자,수행일자
5	색인처리	irflag		F(1)	
6	원문파일명	filename	003	F(50)	
7	권호	volno	004	F(15)	문서번호
8	저자명	author	005	F(50)	발명자,연구책임자
9	국가명	nation	006	F(20)	출원국,개최국
10	잡지명	journal	007	F(30)	
11	출처	source	008	F(50)	연구그룹,연구기관
12	도큐먼트종류	doctype	009	F(15)	자료종류,국제분류
13	제목	title	010	V	회의명,보고서명,기관명
14	주제어	keyword	011	V	
15	요약	abstract	012	V	초록,청구범위
16	내용	content	013	V	본문

로 구성되며, 필드는 각 데이터베이스 특성에 따른 필드로 구성한다.

이 논문에서는 산업/기술 데이터베이스와 논문/잡지 데이터베이스에 대해 구조적 분류기법을 적용하는 것으로 범위를 한정하였다. 이를 위해 16개의 필드 중에서 공통 필드에 있는 레코드 번호, 검색 횟수, 로딩 일자, 국가명, 잡지명, 제목, 주제어, 요약 등 8개의 필드를 기반으로 하여, 정보가 갖는 전형적인 특성에 대한 인덱스 테이블을 구성한다. 또한, 다중 뷰를 구성하는데 사용되는 시간적 정보 및 정보간의 관계에 의한 뷰 간의 동기화를 지원하기 위한 비정형적인 특성에 대한 인덱스 테이블을 추가로 구성한다.

3.3 인덱스 메카니즘

효율적인 정보 시각화를 위해 구조적 정보 분류 기법

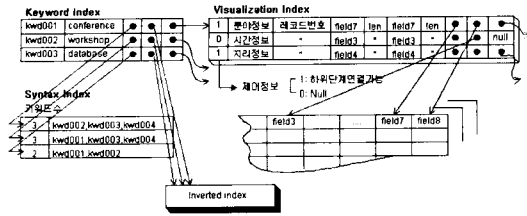


그림 2 색인 구조

을 기반으로 하여, 공통 항목 이외의 특성과 관련성에 대한 색인 구조를 그림 2와 같이 구성하였다. 이를 위해 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. 정보가 갖는 형식화된 공통 특성에 대한 색인 테이블은 레코드 번호, 검색 횟수, 로딩 일자, 국가명, 잡지명, 제목, 주제어, 요약 등 8개 항목으로 구성하였다. 또한 다중 뷰를 구성하는데 사용되는 시간적 정보 및 정보의 영역을 표현하는 뷰 간의 관련성에 의한 이동이 가능하도록, 동기화를 지원하기 위해 동적 테이블을 구성한다. 색인 파일은 저장된 정보의 주제어와 정보 특성을 나타내는 레코드 및 레코드와 관련된 위치 정보인 포인터를 갖는다. 뷰간의 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. 다음은 주제어, 회의명, 개최국 간의 관련성을 테이블 모델, SQL 코드로 나타낸 것이다. 각기 독립적으로 존재하는 색인 항목에 대해, 주제어-ID에 대해 발표된 회의명과 개최국 간의 관련성을 정의함으로써, 다중 뷰를 구성하는 기본 요소를 설정하게 된다. 이러한 관계 정의에 의해 뷰간의 동기화가 지원된다.

Table Model

속성	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. 정보 특성 시각화 시스템

4.1 시스템 구조

전형적인 정보 검색 기법이 데이터베이스로부터 사용자가 원하는 정보를 찾는 데 목적을 둔 사용자 중심이라면, 정보 특성 시각화는 데이터베이스에 무엇이 있는지 보기 위한 것이 목적으로써, 자료 중심 기법이다.

정보 특성 시각화 계층은 정보 검색을 위한 효율적인 정보를 다중 데이터 자원(multiple data source)으로부터 속성 값 쌍(attribute-value pairs)으로 추출한다. 이때, 속성 값의 쌍인 데이터는 각 뷰에 대한 모델에 의해 변환되고, 각 뷰를 통해 상호 작용이 가능하도록 시각화된다. 뷰 계층에서는 뷰를 통해 검색 결과에 대한 아웃라이닝을 사용자에게 보여주고 사용자는 뷰를 통해 상호 작용한다. 전체 시스템의 내부 구조는 그림 3과 같다.

4.2 시스템 구성 요소

정보 특성 시각화 시스템은 검색 결과를 시각적으로 표현하고 표현된 결과로부터 재질의 할 수 있는 뷰, 통합 데이터베이스로부터 각 뷰에 해당하는 자료를 추출하여 저장해 두기 위한 정보 특성 시각화 엔진, 그리고 검색 엔진을 통하여 생성된 정보 특성 시각화 인덱스와 각 뷰를 연결하는 연결자(Mapper)로 나누어 볼 수 있다.

1) 뷰(View)

뷰는 데이터의 특정 속성을 시각적으로 표현한 것으로, 뷰 객체(그래프의 막대 객체, 지도의 특정 위치, 아이콘...), 이들에 대한 정렬, 그리고 객체와 연결된 그래

픽 객체들의 집합으로 구성된다. 사용자는 원하는 정보를 검색하기 위해 뷰를 이용하여 쉽고 편리하게 질의를 할 수 있다. 또한 질의 결과는 데이터의 속성에 따라 각 뷰에 시각적으로 표현된다.

이러한 뷰들은 사용자의 필요에 따라 다양한 형태로 표현할 수 있지만, 이 논문에서는 지리 정보 뷰, 시간 정보 뷰, 그리고 분야 정보 뷰로 한정한다.

2) 뷰 관리자(View manager)

정보가 갖는 특성을 시각적으로 표현하기 위해 그래픽 객체의 집합으로 구성된 뷰 관리자는 키워드 검색에 의한 검색 결과를 각 뷰에 표현하는 일을 담당하고, 각 뷰로부터의 질의를 연결자(mapper)에게 전달하며, 한 뷰로부터의 질의 결과가 다른 뷰에 반영되도록 하는 동기화 및 사용자가 각 뷰를 향해하도록 뷰 간의 이동을 관리한다.

3) 정보 특성 시각화 엔진

여기에서는 그림 3의 속성 모델을 생성하는 부분으로 각 자료를 분류해 놓은 통합 데이터베이스로부터 정보를 추출하여 그 특성에 따라 정보 특성 시각화 인덱스를 생성한다. 추출된 정보의 특성은 뷰의 성격에 따라 달라질 수 있다. 뷰 데이터 관리자(View Data Manager)는 인덱스를 생성하고 뷰 관리자에 의해 변환되어 들어온 질의를 처리하여 그 결과를 다시 각 뷰에 돌려주는 역할을 한다.

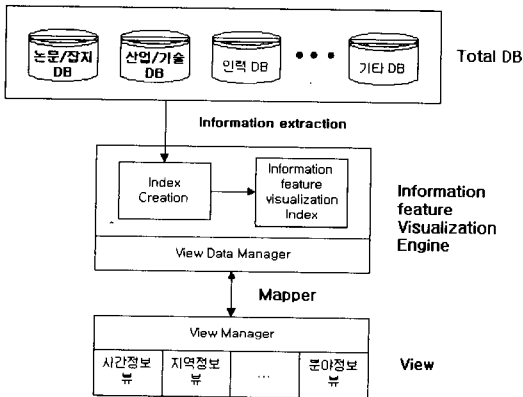


그림 3 시스템의 내부 구조

4) 연결자(Mapper)

연결자(mapper)는 뷰의 물리적인 표현을 관리하는 뷰 관리자와 시각화 인덱스 데이터베이스에 있는 데이터간의 매핑을 위한 수행하는 것으로 표2의 규칙에 따라 관련 정보를 뷰 객체로 연결한다.

표 2 매핑 테이블

뷰	뷰 객체	속성
지리정보뷰	Bar(width)	Count(articles) datei
시간정보뷰	Bar(height) Bar(position)	Count(articles) Locationi
분야정보뷰	Bar(width)	Count(articles) Categoryi

예를 들어, 시간 정보 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) 속성에 따라 표현된 지역의 정보를 나타내며, 분야 정보 뷰의 경우엔 분야(Category) 속성에 의한 정보의 크기를 나타낸다.

5. UML을 이용한 시스템 모델링

여기에서는 UML을 이용하여 정보 특성 시각화 시스템에 대한 모델링 과정을 기술한다[12][13]. 먼저 시스템의 작동에 대한 CRC-Card를 작성하여 이를 기반으로 Class 다이어그램, Sequence 다이어그램, Collaboration 다이어그램 그리고 Activity 다이어그램을 작성한다.

5.1 CRC-Card

정보 특성 시각화 시스템의 동작 과정으로부터 User, Keyword window, IFV, 그리고 View Manager 클래스를 추출할 수 있다. 이들 클래스는 각각 수행해야 할 responsibility를 가지며 다른 클래스들과 상관 관계(Collaboration)를 형성한다. 예를 들어 User 클래스는

표 3 정보 특성 시각화 시스템의 CRC-Card

User	
Keyword를 입력한다.	• Keyword
View내의 component를 이용하여 질의한다	Window
Keyword window	
입력된 Keyword를 전송한다.	• View
선택된 component를 전송한다.	Manager
View Data Manager	
정보를 추출한다	• user
인덱스를 생성한다.	• View
정보 특성 시각화 인덱스를 생성한다	Manager
View Manager	
Mapping을 수행한다	• Keyword
View를 Display 한다	Window
Query를 transform 한다	• View Data Manager

Keyword window를 통해 질의를 위해 입력된 Keyword를 View Manager에게 전달하고, View Manager는 View Data Manager를 통해 색인 정보로부터 요청된 정보와 일치하는 정보를 매핑시킨다. View Data Manager에서는 요청된 정보를 검색하여 그 결과를 View Manager에게 돌려주고 사용자는 View를 통해 결과를 확인할 수 있다. 이것을 CRC-Card로 작성한 결과는 표 3과 같다.

5.2 Class 다이어그램

정보 특성 시각화 시스템의 CRC-Card를 기반으로 추출한 Class들을 표현한 Class 다이어그램은 그림 4와 같다.

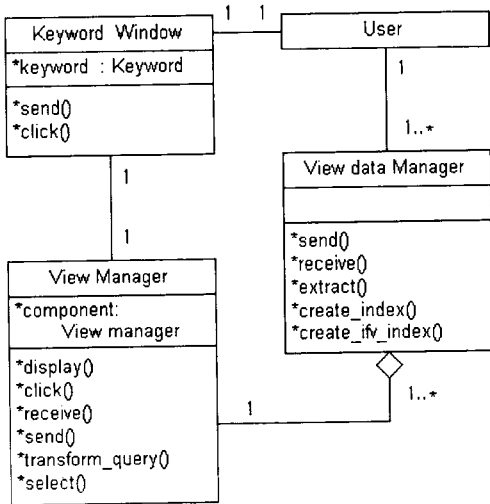


그림 4 class 다이어그램

사용자는 Keyword window Class와 1 대 1 관계를 가지며, 사용자는 한 개 이상의 View data manager를 통하여 질의 및 질의 및 질의 결과를 확인할 수 있고 View data manager 통한 질의는 View manager에 의하여 관리된다. View manager는 View 내의 component를 선택한 확장된 질의를 처리하여 View data manager에게 넘겨주는 중계 역할도 한다.

■ Keyword Window Class

- attribute
 - keyword
- operation
 - send()
 - click()

■ View data manager Class

- attribute
- operation
 - send()
 - receive()
 - extract()
 - create_index()
 - create_ifv_index()

■ View Manager Class

- attribute
 - component
- operation
 - display()
 - click()
 - receive()
 - send()
 - select()
 - transform()

5.3 Sequence 다이어그램

Sequence 다이어그램은 시간의 추이에 따른 객체들 사이의 상호작용을 표현한 것으로 여기에서는 user의 키워드 검색과정의 Sequence 다이어그램과 View를 통한 검색과정의 sequence 다이어그램을 기술한다.

1) Keyword 검색 sequence 다이어그램

사용자로부터 초기 키워드를 입력받아 view를 생성하는 제어의 흐름을 명세화한 것이다. user로부터 Keyword를 Keyword window 객체를 통하여 입력받아 View data manager 객체로 보내면 View data manager 객체에서는 정보를 추출하여 정보 특성 시각화 인덱스를 생성한다. View manager에서는 추출된 시각화 정보의 통계를 구하고 정보를 각 View에 Display한다. 사용자는 표시된 View를 선택하여 다음 질의를 구성할 수 있다. 그림 5는 Keyword 검색을 위한 Sequence 다이어그램이다.

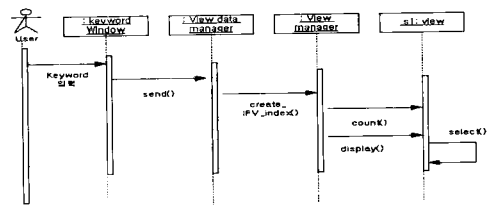


그림 5 Keyword 검색을 위한 Sequence 다이어그램

2) Component 선택을 통한 검색 sequence 다이어그램

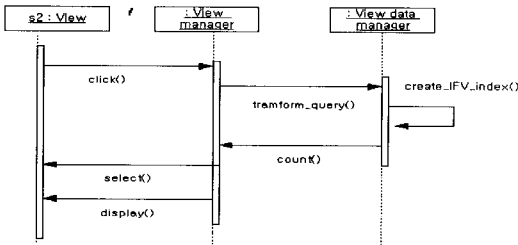


그림 6 component 선택을 이용한 검색 Sequence 다이어그램

표시된 View내의 각 component을 선택하여 질의를 확장한 검색방법의 제어흐름을 명세화 한 것이다. View 객체에서 component를 click하여 view manager 객체에 전달하면 view manager 객체에서는 query를 transform하여 view data manager 객체로 전달하고 이곳에서 정보 특성 시각화 인덱스(IFV_index)가 구성되어 추출된 정보의 통계치를 다시 view manager에게 전달한다. view manager 객체에서는 표시할 view를 선택하고 해당 view에 각 component의 값을 갱신하여 표시한다. 그림 6은 component 선택을 이용한 검색 Sequence 다이어그램이다.

5.4 Collaboration 다이어그램

Collaboration 다이어그램은 객체들 사이의 정적인 구조에 역점을 두어 교환되는 메시지를 표현한 것으로 정보 특성 시각화 시스템의 Collaboration 다이어그램에는 4개의 객체와 각 객체마다 다수의 메시지가 포함되며 그림 7과 같이 표현된다.

- Keyword window 객체
 - keyword 입력 메시지
- View data manager 객체
 - extract()
 - create_index()
 - create_IFV_index()

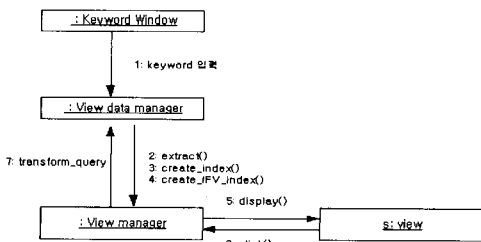


그림 7 collaboration 다이어그램

- View manager 객체
 - display()
 - click()
 - transform_index()

5.5 Activity 다이어그램

Activity 다이어그램은 시스템의 동적인 부분에 대한 모델링 과정으로 정보 특성 시각화 시스템에서는 swimlane을 사용한 workflow 모델링 방법을 적용한다. workflow 모델의 swimlane은 실제세계에 있는 실체를 대표하는 것으로 전체 활동 중에서 각 부분별로 고수준의 책임을 담당한 후 Class로 구현된다. swimlane은 Keyword window, IFV Engine, view manager로 구성되며 그림 8과 같이 표현된다.

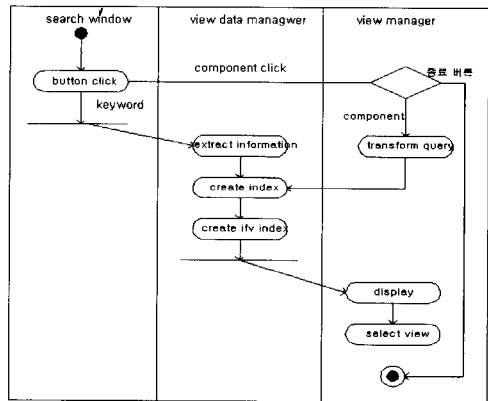


그림 8 Activity 다이어그램

6. 구현

이 시스템은 웹 환경에서 동작하며, C언어를 사용하여 구현하였고, 시스템의 데이터 오브젝트는 통합정보검색시스템에서 사용하는 통합 데이터베이스 스키마를 참조하여 작성되었다. 인포믹스DB로 작성된 테이블로부터 자료를 검색하기 위한 질의 처리는 fulcrum의 SearchSQL를 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 정보뷰, 시간 정보뷰, 분야 정보뷰의 내용도 함께 출력하게 된다. 뷰를 이용한 추가 질의가 주어지면 검색된 정보내에서 새로운 질의 결과와 각각의 뷰들이 재구성되어 출력된다.

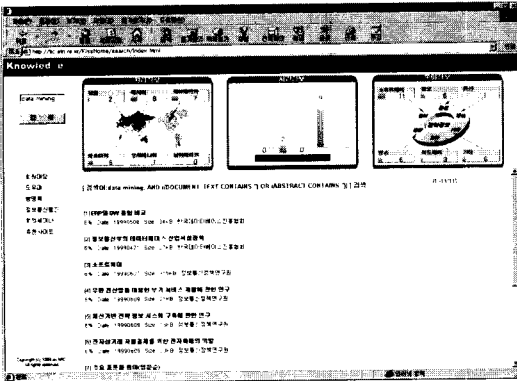


그림 9 정보 특성 시각화 시스템

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, 분야 정보만을 우선적으로 뷰로 구성하였다.

그림 9는 정보 특성 시각화 시스템을 통해 검색한 결과 화면이다.

7. 평가 및 결론

인터넷의 확산으로 보다 많은 정보가 웹 사용자들에게 제공됨에 따라, 보다 쉽게 적합한 정보를 찾기 위한 웹 정보 검색의 필요성이 증가되고 있다. 웹 검색을 위해서는 정보에 대한 분류 방법이 매우 중요하다. 또한 분류된 정보를 시각화함으로써 관련 정보를 파악이 용이하며 효율적인 검색이 가능하다. 이 논문에서는 정보 영역과 관련성을 기반으로 한 구조적 정보 분류 기법을 제안하였고, 각 시스템의 기능과 논리적인 정보 표현을 위한 색인 구조를 설계하였다. 또한, UML을 이용하여 이러한 분류 구조를 기반으로 한 정보 특성 시각화 시스템을 모델링 하였다.

기존의 검색 방법은 단순히 주제를 통한 검색 결과에 대한 목록을 나열하므로, 사용자가 원하는 정보를 파악하기 위해서는 입력된 주제어에 의해 생성된 1차 정보 집합인 검색된 목록을 모두 확인하여야 할 뿐 아니라, 정보의 관련성 및 동향을 파악하기 어렵다. 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 다음과 같은 기대 효과를 얻을 수 있다. 첫째, 사용자들은 주제어를 통한 검색뿐 아니라 정보의 특성에 따라 만들어진 다양한 뷰를 이용하여 검색을 수행하므로 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 관리가 용이하다. 왜냐하면, 정보 영역에 있는 항목이 변경되더라도 관계형 모델은 영향을 받지 않기 때문에, 수시로 변하는 웹 정보를 관리하는데 매우 효과적이다.

향후 연구로는 의미를 기반으로 한 정보 추출 및 클러스터링 기법에 대한 연구와 3차원을 통한 정보 시각화와 멀티미디어 검색에 대한 연구가 요구된다.

참고 문헌

- [1] Huser C, Reichenberger K, Rostek L, Streitz N, "Knowledge-based editing and visualization for hypermedia encyclopedias," Communications of the ACM, Vol. 38, No. 4, April, 1995, pp. 49-51.
- [2] Fox EA, Hix D, Nowell LT, Brueni DJ, Wake WC, Health IS, Rao D. Users, "user interface and objects : Envision, a digital library," Journal of the American Society for Information Science, Vol. 44, No. 8, September, 1993, pp. 480-491.
- [3] 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.csd.tamu.edu/DL95/papers/nuernberg.html>.
- [4] Romana Rao, Jan O. Pederson, Marti A. Hearst, etc. "Rich Interaction in the Digital Library," Communications of the ACM, Vol. 38, No. 4, April, 1995, pp.29-39.
- [5] Hearst, M., TileBars: Visualization of Term Distribution Information in Full Text Information Access, Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, Denver, CO, 1995.
- [6] 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.
- [7] Rao R, Card SK, Jelinek HD, Mackinlay 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] Koichi Takeda and Hiroshi Nomiyama. "Information Outlining for Digital Libraries," Research Report-RT0181, IBM Research, Tokyo Research Laboratory, Jan., 1997.
- [9] Comp, Introduction to the CR Classification System, Prentice-Hall, 1985.
- [10] R.Prieto-Diaz and P. Freeman, "Classifying Software for Reusability," IEEE Software, Vol.4, No.1, Jan., 1987, pp.6-16.
- [11] T. J. Lee, "Software Reuse in Parallel Programming Environments," Ph.D. dissertation, Department of Computer Science, The University of Texas at Austin, 1989.
- [12] G. Booch, J. Rumbaugh, I. Jacobson. "The Unified Modeling Language User Guide," Addison Wesley Longman, Inc. February, 1999.
- [13] Craig Larman, "Applying UML and Patterns," Prentice Hall. 1998.



조 윤 기

1994년 충북대학교 자연대학 컴퓨터과학과 졸업(이학사). 1996년 충북대학교 자연대학 전자계산학과 석사(이학석사). 1996년 ~ 1997년 현대정보기술 기술연구소 연구원. 1998년 ~ 현재 충북대학교 전자계산학과 박사과정. 관심분야는

객체지향 모델링, 정보 시각화, 정보 검색.



하 재 관

1993년 대전대학교 전자계산학과 졸업(공학사). 1996년 청주대학교 전자계산학과 석사(공학석사). 1999년 8월 충북대학교 대학원 박사과정 수료. 1996년 ~ 현재 우송공업대학 무역사무자동화과 조교수. 관심분야는 객체지향 테스팅, 품질관리,

정보 시각화.



구 연 설

1964년 청주대학교 졸업. 1975년 성균관대학교 경영행정대학원 전자자료처리 전공(경영학석사). 1981년 동국대학교 대학원 통계학 전공(이학석사). 1988년 평운대학교 대학원 전자계산학 전공(이학박사). 1994년 ~ 1995년 한국정보과학회

부회장. 1979년 ~ 현재 충북대학교 컴퓨터과학과 교수. 관심분야는 객체지향 테스팅, 품질관리, 정보 검색, 전자 도서관.