

# WWW상에서의 계층적 구조를 이용한 정보 시각화 도구 개발

노 원 빈<sup>†</sup> · 허 신<sup>†</sup>

## 요 약

WWW(World Wide Web)에서는 하이퍼텍스트를 기반으로 문서들이 서로 연결되어 있다. 그러므로 큰문서 공간을 검색하고자 할 때, 웹 사용자는 공간 내에서의 자신의 위치를 기억해야 하고 다음에 갈 곳을 결정해야 하며, 전에 방문한 페이지의 위치를 유지해야만 한다. 이러한 일로 인해, 하이퍼텍스트 검색 작업은 방향 상실의 문제점을 가지고 있다. 그러므로 웹에서의 정보를 시각화하는 일은 필요한 일이며, 서로 연결된 문서들의 집합에서 정보의 양과 분포를 표현하는 일은 어려운 문제이다. 문서의 형태의 따라 시각화한 정보들(문서, 그림, 음악, 기타)은 사용자가 깊고 복잡한 구조를 브라우징하는데 걸리는 많은 시간을 줄여준다. 본 논문에서는 이러한 문제점을 해결하기 위하여 계층적 구조를 이용한 웹 정보 시각화 도구인 WIV(Web Information Viewer)를 개발하였다. WIV에서는 네 가지 전략 즉 바인딩, 클러스터링, 필터링, 계층화를 사용하여 효과적인 개요를 제공한다. 또한 자바(Java)로 구현되어졌기 때문에, 하드웨어 독립적으로 사용되어질 수 있고 자바를 지원하는 모든 웹 브라우저에서 사용할 수 있도록 설계하였다.

## Development of a Tool for Information Visualization Using Hierarchical Structure on the WWW

WonBin Rho<sup>†</sup> · Shin Heu<sup>†</sup>

### ABSTRACT

Hypertext-based documents are linked together on the World Wide Web(WWW). When navigating thru large documents, hypertext readers must remember their locations in the space, make decisions on where to go next, and keep track of the pages previously visited. This can cause problems such as evidence of disorientation in a task of hypertext-based search. Therefore, it is needed to visualize the information on the WWW such that the problem of information visualization is to represent the quantity and distribution of information within a set of linked documents. This visualized information along with the type of document i. e., text, image, audio, etc., can be helpful when deciding whether a web site may be desired, without spending a great deal of time to browse the deep and complex structure of the site. In this paper, we develop a tool for web information visualization, namely Web Information Viewer(WIV), using a hierarchical structure to solve the problems mentioned above. It consists of four strategies: binding, clustering, filtering, and hierarchization to show the effectiveness. In addition, the design provides hardware-independency of all browsers that can support Java.

---

<sup>†</sup> 정 회 원: 한양대학교 전자계산학과  
논문접수: 1997년 11월 26일, 심사완료: 1997년 12월 15일

## 1. 서 론

최근에 인터넷(Internet)에 대한 관심이 높아지고 있다. 그 이유는 아마도 정보가 재산이 되는 정보화 사회에서 가장 많은 정보를 손쉽게 제공해 주기 때문일 것이다. 이러한 인터넷에서 제공되는 서비스 중에서 월드 와이드 웹(WWW, World Wide Web)은 사용하기 편리한 그래픽 사용자 인터페이스(GUI)와 멀티미디어 환경을 제공하기 때문에, 사용자 수가 폭발적으로 늘어나고 있는 실정이다.

하지만 WWW(World Wide Web)상에서 사용자가 큰문서 공간(large document space)을 항해하는(navigate) 하는 일은 어려운 일이다. 왜냐하면, 종종 방향 상실(disorientation)에 대한 문제를 경험하기 때문이다[10][12]. 한 위치에서 다른 위치로 이동하는 과정은 사용자에게 쉽게 혼란을 가져다 줄 수 있다. 그래서 웹 사용자(하이퍼텍스트 사용자)는 망에서 자신의 위치를 기억해야 하고 다음에 갈 곳을 결정해야 하며, 전에 방문한 페이지를 유지해야만 한다. 이러한 일로 인해, 하이퍼텍스트 검색 작업은 방향 상실의 문제점을 가지고 있다.

이러한 방향 상실의 문제점은 정보 공간의 전체적인 구조에 대한 사용자의 지식 부족의 결과라고 말할 수 있다. 그러므로 하이퍼텍스트 설계를 향상시키는 방법 중 한 가지는 사용자에게 구조적인 개요(overview)를 제공해 주는 것이다. 하이퍼텍스트 내용에 대한 구조적인 개요를 제공하는 것은 사용자의 방향성을 촉진할 수 있다. 즉 방향 상실의 문제점을 해결할 수 있다. 하지만 효율적인 개요를 구성하는 것은 여러 가지 이유로 도전이 되는 일이다. 정보 공간의 내용과 구조의 지식이 없이 효율적인 개요를 발견하는 것은 어려운 일이고 개요를 발견하여도 웹 구조를 그래프로 표현하는 것은 매우 복잡한 문제이다. 또한 하부의 정보 공간의 크기가 증가함에 따라 전체적인 구조를 한 화면에 표시하는 일은 매우 어려운 일이다. 이러한 경향은 전체적인 구조를 분명치 않게 만들 수 있고 검색의 어려움을 줄 수 있다. 그러므로 웹에서의 정보를 시각화하는 일이 필요한 일인데, 서로 연결된 문서들의 집합에서 정보의 양과 분포를 표현하는 일은 어려운 문제이다. 문서의 형태의 따라 시각화한 정보들(문서, 그림, 음악, 기타)은 사용자가

깊고 복잡한 구조를 브라우징하는데 걸리는 많은 시간을 줄여준다.

웹 상에서의 정보 시각화에 대한 연구는 많이 수행되어져 왔다. 웹에서의 방향 상실의 문제점을 해결한 다양한 도구들이 개발되어져 왔는데, 그 중에서는 브라우저나 렌즈와 같은 항해 도구, 히스토리 리스트나 책갈피, 필터와 같은 수동적인 프로세스(passive process), 에이전트와 같은 능동적인 프로세스(active process) 등이 있다. 몇 가지 예로서, Pittsburgh 대학에서 개발한 정보 시각화 도구인 CASCADE(Computer Augmented Support for Collaborative Authoring and Document Editing)는 문서들의 협력 작업을 지원하는 항해 도구와 수동적, 능동적 프로세스들을 하나로 통합하여 계층적 구조의 개요를 통한 정보 검색을 위한 시각적 인터페이스를 제공해준다[11]. WebBook은 사용자가 웹 활동에 대해서 바인딩(binding) 작업을 통해서 조직적으로 체계화할 수 있도록 구성되어있다[3]. 또한 Navigational View Builder는 정보 공간의 유용한 시각화를 위하여 필터링(filtering)과 클러스터링(clustering)으로 효과적인 개요를 제공해주고 있다[9].

이 외에도 웹 상에서의 정보 시각화에 대한 연구가 끊임없이 수행되고 있으며 사용자를 고려한 많은 도구들이 개발되고 있다[2][6][7][8]. 하지만 위에서 언급한 도구들은 모두 플랫폼 호환성이 없다. 즉 현재 웹 서버와 클라이언트를 위한 환경은 참으로 다양한데, 그러한 환경에서 모두 호환성 있게 수행될 수 있는 도구가 필요하다. 또한 위의 도구들은 웹 정보에 대한 효과적인 시각화 정보를 제공해 주지만 그 정보를 기존의 브라우저와 같은 도구에서 재 사용할 수 없다는 단점을 가지고 있다.

또한 웹 정보를 시각화하기 위해서 그 정보를 한 눈에 알아볼 수 있는 개요와 정보들간의 관계를 표현할 수 있는 인터페이스(interface)에 대한 선택이 중요하다. 보통 목차(contents)는 서적에서 개요나 항해를 제공하는 일반적인 방법으로 알려졌다. 또한 큰 계층적인 목차를 검색하기 위한 인터페이스로는 확대/축소 인터페이스가 있는데, 전체적으로 확대된 계층적 구조를 보여주는 것에 비해서 브라우징 시간을 줄여준다[5]. 문서들의 정보를 표현하는 인터페이스로는 주제들을 위한 공간을 제공해 주는 전통적인 들여쓰기 리스트 형식(indented list format)과 크기의 비교를

가능하게 하는 2차원적인 길이 막대 그래프가 있다. 웹을 시각화하는 작업에서 3차원이나 모든 연결을 표현하는 그래프가 사용되어져 왔지만, 이러한 시각화 작업은 전형적으로 목차를 사용하고 배우는 것보다 복잡하다[1]. 그러므로 2차원적인 길이 막대 그래프는 큰 정보 공간을 함해하고 시각화하는데 사용자가 항목의 속성 값을 쉽게 알아 볼 수 있게 해준다[4].

본 논문에서 개발한 WIV(Web Information Viewer)는 방향 상실과 개요 구성시 어려움 등을 계층적인 개요를 생성해줌으로서 해결할 수 있다. 기존의 도구들의 단점들을 보완하고 장점들을 통합한 네 가지 전략 즉 정보의 특성과 시각적 특성을 바인딩(binding)하는 작업, 확대/축소 기능으로 많은 정보를 효율적으로 보여주는 클러스터링(clustering), 불필요한 정보에 대한 필터링(filtering), 개요의 계층화(hierarchization)를 통해서 이러한 문제들을 해결할 수 있다. 또한 개발에 사용한 언어는 보안 기능이 있다는 점과 플랫폼 독립적인 특성으로 이식이 쉽다는 점 그리고 웹을 위한 프로그래밍 언어라는 점에서 자바(Java)를 사용하였다[14].

본 논문의 구성은 다음과 같다. 2장은 본 논문에서 개발한 도구인 WIV(Web Information Viewer)의 구조와 네 가지 전략에 대해서 자세하게 기술한다. 3장에서는 이러한 WIV의 구현 결과와 활용 방안을 살펴보고 4장에서 결론 및 향후 과제를 기술한다.

## 2. 웹 정보 시각화 도구: WIV

### 2.1 WIV의 개요

본 논문에서 개발한 웹 정보 시각화 도구인 WIV는 사용자가 큰문서 공간의 웹 사이트의 정보들을 계층적 구조인 개요(overview)로 시각적으로 보여주는 도구이다. 하이퍼텍스트 내용에 대한 계층적인 구조의 개요를 제공하는 것은 사용자의 방향성을 촉진할 수 있다. 즉 방향 상실의 문제점을 해결할 수 있다. 기존의 정보 시각화 도구들은 플랫폼 호환성이 없고 제공한 시각화 정보를 재사용할 수 없었지만, 본 논문에서 개발한 WIV는 그러한 단점들을 보완하였고, 기존의 도구 각각의 장점들을 고려하여, 바인딩, 클러스터링, 필터링, 계층화와 같은 4가지 전략을 사용하여 문제점들을 해결한다. WIV는 다음과 같은 기능을 수행

하도록 설계되어졌다.

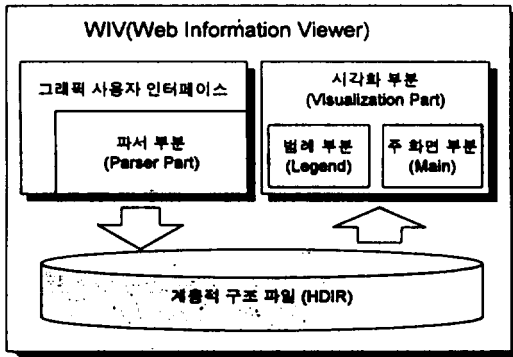
- 원하는 웹 사이트를 파싱(parsing)하여 계층적인 구조의 개요로 만들어 준다. 이러한 것은 두 가지 방법으로 생성할 수 있는데, 존재하는 연결(link)에 따른 방법과 디렉토리와 파일 구조를 사용하는 방법이다.
- 파싱된 계층 구조를 파일로 저장하여 재사용이 가능하다. 이 파일은 확장자가 \*.HDIR로서 원하는 웹 사이트에 대한 여러 가지 정보를 가지고 있고 계층적으로 표현되어 있다.
- 웹 사이트에 대한 정보들(문서, 그림, 음악, 기타 등등)을 서로 다른 색으로 구분하여 표시하였다.
- 전통적인 들여 쓰기 리스트 형식(conventional indented list format)에서 주제를 위한 공간을 제공해 주고 단일 차원 선과 함께 크기의 비교를 가능하게 해준다. 즉 각 정보들의 항목 수와 크기를 알 수 있도록 길이 막대 그래프(value bar)를 사용하여 표시하였다.
- 주제별, 그래프별, 주제와 그래프별로 정보를 다양하게 시각화 할 수 있다.
- 내부의 연결과 외부의 연결을 선택할 수 있다. 이러한 기능을 제공해주므로 그 웹 사이트 내부의 실질적인 크기를 대략적으로 알 수 있고 외부로의 연결에 대한 정보를 알 수 있다.
- 제공된 개요에서 원하는 항목을 선택 시 그 문서로 바로 이동할 수 있다. 이러한 기능은 원하는 정보를 얻기 위해 처음부터 찾아야 하는 불편함을 없앴다.
- 찾고자 하는 단어나 구를 검색(search)할 수 있는 기능이 제공된다. 검색하는 단어나 구를 포함한 주제들이 모두 시각적으로 표시되므로 원하는 정보 쉽게 찾을 수 있다.
- 계층적 구조를 사용하여 레벨(level)을 사용자가 확대(expand)하고 축소(contract)하는 기능을 제공해주므로 전체적으로 확대된 계층적 구조를 보여주는 것에 비해서 브라우징 시간을 줄여 준다.
- 자바로 구현되어졌기 때문에, 하드웨어 독립적으로 실행할 수 있다.

이러한 기능을 가진 웹 정보 시각화 도구인 WIV는

파일이나 페이지들이 어디에 위치한지를 빨리 보여 주고 디렉토리들의 상대적인 크기들을 비교하고 평소에 잘 찾지 않는 계층적 구조의 큰 가치를 보기 위해 웹 사이트와 디렉토리를 브라우징하는 것을 도와 줄 것이다.

2.2 WIV의 전체적인 구조

WIV 시스템의 구조는 크게 두 부분으로 나눌 수가 있다. 파서 부분과 시각화 부분이다. 파서 부분은 웹 사이트를 파싱하여 계층적인 파일 구조로 만들어 주는 부분이고 시각화 부분은 이 파일을 사용하여 원하는 웹 사이트에 대한 정보를 시각적으로 보여주는 부분이다. (그림 1)은 WIV의 전체적인 구조에 대해서 보여주고 있다.



(그림 1) WIV의 전체적인 구조  
(Fig. 1) The overall structure of WIV

2.3 자료 구조

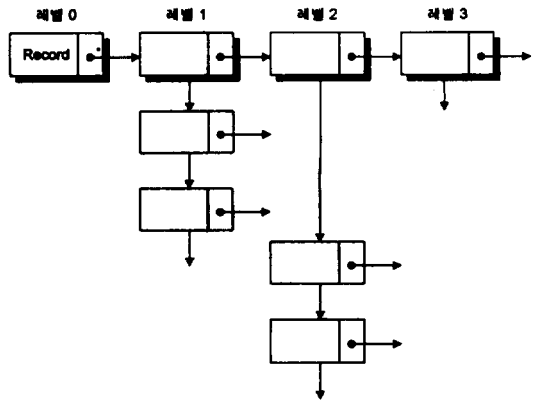
본 논문에서 개발한 웹 정보 시각화 도구인 WIV는 웹 사이트에 대한 정보를 계층적인 구조로 만들어주기 위한 자료 구조를 가지고 있다. 파서 부분은 일단 문서를 파싱하여 본 시스템에서 원하는 자료 구조로 바꾸어준다. 이러한 자료 구조를 바탕으로 계층적으로 확대해서 파싱을 하고 마지막으로 계층적인 파일을 생성한다. 시각화 부분에서는 이러한 계층적 구조 파일을 사용하여 다시 시스템에서 필요한 자료 구조로 바꾸어주고 화면에 시각적으로 표시해준다.

WIV는 파서 부분과 시각화 부분에서 한 개의 웹 페이지에 대한 정보를 일시적으로 저장하기 위해서

다음과 같이 Record 클래스를 정의하였다.

```
class Record {
    public double fileSize; // 파일의 크기
    public String hrefadd; // 하이퍼텍스트 참조
                          //에 첨가될 태그
    public String href; // 하이퍼텍스트 참조
    public String title; // 주제
    public String type; // 파일의 종류
    public Vector recList; // 연결 리스트
    public int index; // 처음 색인
    public double imgtotal; // 총 그림 개수
    public static int recordCount = 0;
    public int toplevel = 1000;
```

Record 클래스는 하나의 웹 페이지에 대한 정보를 가지고 있다. 파일의 크기와 하이퍼 참조의 절대 경로, 정보의 종류, 레코드 개수, 등 시각화를 위해 필요한 정보들을 포함하고 있다. 이러한 레코드들이 서로 연결 리스트(linked list)로서 계층적인 구조를 이루게 된다. 이러한 자료 구조 형태를 (그림 2)에서 보여주고 있다.



(그림 2) WIV의 계층적 자료 구조  
(Fig. 2) Hierarchical data structure of WIV

(그림 2)에서 보여주는 것처럼, 레벨(level)을 두어서 레벨의 맨 처음에 시작하는 레코드들만 연결 리스트로 구성한다. 이 연결 리스트는 클래스 Vector를 사

용하여 구성할 수 있다[13][15]. 그리고 각각의 레벨에서 또 다른 연결 리스트를 구성함으로써 계층적 구조 형태를 지니게 된다. 그러므로 웹 사이트에 대한 정보를 계층적으로 표현할 수 있고 파싱이 가능하게 된다.

2.4 파싱 방법

문서에 포함된 연결(link)을 다른 웹 페이지에 파싱은 ConvertFile 함수에서 일단 HTML문서의 다음과 같은 태그(tag)를 찾아서 수행하게 된다.

- <FRAME SRC="....." .....
- <IMG SRC="....." .....
- <A HREF="....." .....
- <AREA .....
- <ISMAP SRC="....." .....
- <HL>.....</HL>

이 가운데서 <HL> 태그는 본 논문에서 새롭게 정의한 태그이다. 자세한 사항은 WIV의 전략 중 필터링(filtering) 부분에서 기술한다. HTML문서의 이러한 태그를 찾아 파싱을 하게 되는 이렇게 하이퍼 참조된 문서를 파싱하기 위해서 본 논문에서 사용한 알고리즘은 폭 우선 확장(Breath First Expansion)이다 [9]. 루트(root)에서 시작하여 각 레벨에서 모든 참조된 연결을 찾아서 기록하고 하위 레벨로 넘어가게 된다. 즉 연결(link)이 루트와 가장 가까운 계층적 구조의 레벨에서 표시되어지는 것을 가능하게 하기 위해서 이 알고리즘을 사용한다.

또한 앞서 살펴보았듯이 WIV는 두 가지 파싱 전략을 제공해 준다. 존재하는 연결에 따른 방법과 디렉토리와 파일 구조를 사용하는 방법이다. 연결에 따른 방법은 존재하는 웹 사이트를 위해서 적절한 반면, 디렉토리 구조를 사용하는 방법은 새로 디지털화하는 아직 처리 안된 즉, 아직 연결되거나 색인화되거나 주석화되지 않은 수집물을 위해 적절하게 사용할 수 있는 방법이다.

이렇게 모든 연결을 따라가는 것은 정상적으로 전체 웹 사이트를 포함하게 된다. 하지만 사이트의 모든 경로(path)를 따라가는 것이 항상 가능한 것은 아니다. 예를 들어 검색(search)이나 폼(form)의 다른 종류가 사이트의 부분을 접근하는데 사용되었을 때,

파서(parser)는 단지 폼(form) 페이지만 볼 수 있다. 정상적인 구조에서 생성된 cgi-bin 파일들은 정적인 파일처럼 스캔되어진다.

2.5 WIV의 전략

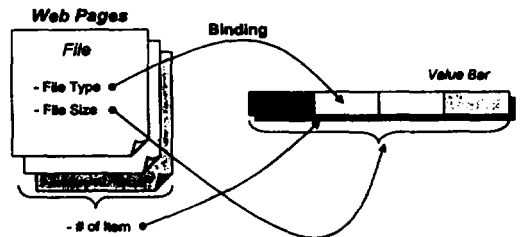
WIV 시스템은 웹 상의 문제점을 해결하기 위해서 정보의 특성과 시각적 특성을 바인딩하는 작업, 확대/축소 기능으로 많은 정보를 효율적으로 보여주는 클러스터링, 불필요한 정보에 대한 필터링, 개요의 계층화와 같은 네 가지 전략을 사용한다. 본 절에서는 네 가지 전략 즉 바인딩(binding), 클러스터링(clustering), 필터링(filtering), 계층화(hierarchization)에 대해서 자세히 기술한다.

(1) 바인딩(Binding)

사용자에게 유용한 개요를 만들어주기 위해서는 정보 공간의 구조뿐만 아니라 보다 많은 정보를 기술해야만 한다. 사용자에게 원하는 정보 공간에 대한 내용에 시각적으로 보여줌으로 사용자가 계층적인 개요만을 봄으로서 전체 정보 공간의 어떤 부분이 자신이 흥미있어 하고 원하는 부분인지를 결정할 수 있도록 한다. 그러므로 시각적인 특성이 하부의 정보 공간으로부터의 정보를 표현하기 위해 계층적인 개요에 사용되어질 수 있다.

계층적인 개요의 설계자는 정보의 속성과 노드나 연결의 시각적인 특성사이의 바인딩(binding)을 명시할 수 있어야만 한다. 이러한 바인딩은 정보 공간과 하이퍼미디어 시스템의 시각적인 공간 사이에 사상(mapping)을 명시하게 된다[3].

본 WIV 시스템에서는 웹 페이지에 대한 정보 특성들과 시각적 특성들 사이의 바인딩을 (그림 3)과 같이



(그림 3) 바인딩(Binding) 전략  
(Fig. 3) Binding strategy

명시하였다. 정보의 특성들은 파일의 종류와 크기, 항목의 수 등이며 시각적 특성으로는 길이 막대 그래프(value bar)를 사용하여 파일의 종류는 색(color)을 사용하여 표현하였고 파일의 크기는 그래프의 길이로, 항목의 수는 그림자로 표현하였다.

## (2) 클러스터링(Clustering)

개요를 제공해 줄 때 생기는 주요한 문제 중 하나는 화면에 너무나 많은 정보가 무질서하게 놓여 있다는 점이다. 이러한 문제를 해결하는 한가지 방법은 적절한 추상화(abstraction)를 통해서 화면상에 보여지는 필요한 정보를 줄이는 것이다. 이러한 작업을 클러스터링(clustering)이라고 한다. 이러한 추상화는 사용자가 화면상에서 무질서하게 놓여지는 일 없이 전체적인 정보 공간을 볼 수 있게 한다. 클러스터링의 두 가지 형태는 다음과 같다[9].

- 구조 기반 클러스터링(structure-based clustering)
  - 구조 기반 클러스터링에서 하이퍼미디어 구조는 주어진 패턴과 일치하는 하위망을 위해 검색되어진다. 이러한 형태의 가장 기본적인 것은 연결의 특별한 클래스에 의해 결합되어진 노드들을 서로 묶어주는 연결 기반 클러스터링(link-based clustering)이다.
- 내용 기반 클러스터링(content-based clustering)
  - 이러한 형태에서는 노드는 개별적으로 고려되어지고 그들은 특성은 클러스터(cluster)를 결정하는 데 검사되어진다. 가장 기본적인 형태는 어떤 특성을 만족하는 모든 노드들을 클러스터로 형성하는 것이다. 또 다른 것으로는 속성 유사성 기반 클러스터링(attribute-similarity-based clustering)이 있는데 이것은 특정한 속성의 각 값에 대한 노드의 클러스터를 형성하는 방법이다.

이러한 클러스터링(clustering)의 형태 중에서 WIV 시스템은 구조 기반 클러스터링(structured-base clustering)을 사용한다. 각 연결에 레벨(level)을 두어서 그 하위 연결들은 하나로 묶는다. 그리고 확대/축소 기능을 사용하여 화면상에 효율적으로 보여지도록 한다.

## (3) 필터링(Filtering)

웹을 통해 들어오는 정보들을 브라우저로 보내기 전에, 불필요한 정보를 삭제하거나 추가적인 내용을 덧붙이는 작업을 정보 필터링(filtering)이라고 한다 [9]. 연결(link)을 따라서 웹 페이지를 파싱하는 작업만으로는 정보를 효율적으로 보여줄 수가 없다. 웹 서버 관리자가 사용자에게 효율적인 개요를 보여주기 위해서 연결에 따른 파싱으로 개요만 보여준다면, 그것은 효과적인 개요가 될 수 없다. 관리자는 웹 페이지의 어떠한 불필요한 정보에 대해서 연결에 따른 파싱을 원치 않을 수도 있다. 또한 체계적이고 효율적인 정보 시각화를 위해 내용을 분류할 수도 있을 것이다. 이러한 필터링 작업을 위해서 본 시스템에서는 HTML문서에 <HL> 이라는 새로운 태그(tag)를 추가하였다. 그 형식을 살펴보면 다음과 같다.

```
<HL [ignore = true/false] [label = "....."]>
<\HL>
```

원하지 않는 정보에 대해서 필터링을 하기 위해서는 다음과 같은 쌍으로서 표시할 수 있다.

```
<HL ignore = true>
원하지 않는 정보에 대한 연결들
<HL ignore = false>
```

또한 좀 더 체계적이고 효율적인 정보를 시각화하기 위해서는 다음과 같은 쌍으로 표시할 수 있다.

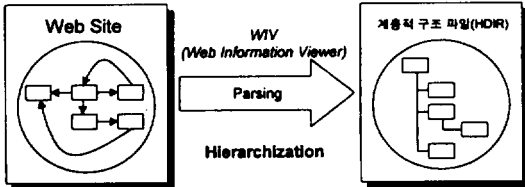
```
<HL label = "서론">
서론에 해당하는 정보들의 연결들
<\HL>
<HL label = "본론">
본론에 해당하는 정보들의 연결들
<\HL>
```

이상과 같이 불필요한 정보를 삭제하고 필요한 정보를 계층적인 개요에 추가하기 위해서 <HL>이라는 새로운 태그를 사용하면 된다. 파서(parser)에서는 이 태그를 인식하여 계층적 구조 파일(HDIR)에 첨가하게 된다. 이 <HL> 태그는 HTML 문서에서는 사용되

지 않는 태그로서 하나의 주석문으로 여기어 인식을 하지 않는다.

**(4) 계층화(Hierarchization)**

하이퍼 문서로서 서로 연결되어 있는 WWW는 비구조적이기 때문에, 연결에 따른 파싱으로 웹 사이트의 개요를 만들면 앞서 살펴본 바와 같이 노드와 링크의 그래프로 표현되는데, 그래프 배치는 매우 복잡한 문제이다. 따라서 이런 복잡하고 비구조적인 웹 사이트를 계층화(hierarchization)할 필요가 있다. (그림 4)는 계층화에 대해서 전반적으로 보여주고 있다.



(그림 4) 계층화 전략  
(Fig. 4) Hierarchization strategy

**1) 중복된 URL의 제거**

앞서 파서 부분에서 원하는 웹 사이트를 파싱하는 폭 우선 확장(breadth first expansion) 방법은 보통의 검색 엔진에서 자주 사용하는 방법이지만, 웹의 구조가 반복적으로 다시 처음으로 되돌아 와서 참조하는 경우도 있듯이 계층적인 구조가 아니므로 이러한 구조를 계층적인 구조로 만들기 위해서는 중복된 URL을 제거할 필요가 있다. 다음과 같은 방법으로 중복된 URL을 제거한다.

- ① 해쉬 테이블(hash table)을 구성한다.  
`ht = new Hashable(20000, 0.5f);`
- ② 각 레코드들을 해쉬 테이블에 넣는다. 키(key) 값은 하이퍼텍스트의 경로(htpath)가 된다.  
`ht.put(htpath, newr);`
- ③ 문서를 파싱할 때, 하이퍼텍스트 참조의 경로를 발견하면, 해쉬 테이블에서 키 값과 비교하여 같은 값이 있으면 파싱하지 않고 제거한다.  
`ht.containsKey(htpath);`

이러한 방법으로 중복된 URL을 제거함으로써 웹 사이트의 비구조적인 것을 계층적인 구조로 만들 수 있다.

**2) 계층적 파일 생성**

WIV의 파서 부분에서는 웹 사이트에 대한 정보를 자동적으로 계층적 구조의 파일로 생성해준다. 파서에 의해서 생긴 이 파일은 URL 기반의 서버와 문서, 연결의 그룹의 시작과 끝의 지시자(indicator) 그리고 크기, 파일 이름, 종류, 내용 등을 포함한 각 문서나 연결에 대한 정보를 포함하는 면에서 HTML 문서와 유사한 형식을 가지고 있다. 파일의 확장자는 ".HDIR"이다. 다음은 계층적 구조 파일(HDIR)의 형식에 대해서 기술한다.

- <HDIR> -파일의 시작
- <SERVER> 서버 URL -서버 URL의 정의
- <BASE> 시작 디렉토리 -시작 디렉토리 위치
- <R> 파일 크기, "URL", 파일 종류, 내용  
 -레벨 0의 헤더
- <HL> 레벨 디렉토리 -레벨의 시작
- <R> 파일 크기, "URL 또는 파일", 파일 종류, "내용"  
 -레벨 1의 항목
- <R> 파일 크기, "URL 또는 파일", 파일 종류, "내용"  
 -레벨 1의 항목
- <R> 파일 크기, "URL 또는 파일", 파일 종류, "내용"  
 -레벨 1의 항목
- </HL> 레벨 디렉토리 -레벨의 끝
- </HDIR> -파일의 끝

**※ 파일 종류**

- h: HTML문서나 문서 파일
- i: 그림 파일
- a: 음악 파일
- o: 기타 파일
- c: CGI 관련

각 계층적 구조에서의 레벨(level)은 <HL> 태그를 사용하여 증명될 수가 있다. 또한 내용 뒤에 선택적으로 첨가되는 숫자는 내부의 그래픽 크기를 나타내 준다. 이런 HDIR 파일은 지역적인 웹 사이트의 문서

로부터 크기나 연결을 표시한다. 서버의 URL은 단지 서버에 지역적인 문서만 따르는 연결로 제한한데 사용되어진다. 즉 외부의 연결을 표시하는 선택사항은 있지만, 문서의 크기와 종류를 지시하는 것은 없으므로 전체 웹 사이트에 대한 문서의 크기에는 이 외부 연결은 영향을 미치지 않는다.

**3. 구현 결과 및 활용 방안**

**3.1 파서 부분**

본 논문에서 개발한 WIV의 파서 부분은 문서에 포함된 연결(link)을 따라 개요를 자동적으로 생성해 주고 계층적 구조의 서로 다른 레벨로서 연결의 새로운 집합인 계층적 구조 파일(HDIR)을 다루게 된다. 파서 부분은 사용자가 독립적으로 실행할 수 있도록 자바 응용 프로그램(application)으로 구현되어졌다. 원하는 웹 사이트에 대한 정보를 얻기 위해서 그 사이트의 URL, 하이퍼문서 참조, 계층적 구조를 위한 최대 레벨 등의 위치(location) 정보가 사용되어진다. (그림 5)는 WIV에서 문서를 파싱하기 위한 위치 정보를 묻는 대화 상자이다. 예를 들어 국회 도서관을 레벨 10으로 하여 파싱하였다.

이러한 위치 정보를 사용하여 그 웹 사이트를 파싱하게 된다. 참조된 연결에 따라 레벨로 파싱을 하게 되는데, 파싱 과정을 마친 후에는 그 웹 사이트에 대한 정보를 계층적 구조로 표현한 HDIR 파일이 생성된다. 이 파일은 시각화 부분에서 웹 정보 시각화를 위해서 사용되어진다. 다음은 국회 도서관을 파싱한 후 생성된 계층적 구조 파일의 일부를 보여주고 있다.

```

(HDIR) 969
(SERVER)http://www.nanet.go.kr
(BASE)/
  (R)3693.0, "/", h, "국회도서관에 오신 것을 환영합니다.", 44271.0
(HL)/
  (R)1603.0, "nal/whatsnew/whatsnew.htm", h, "도서관 소개"
(HL)/nal/whatsnew/
  (R)15851.0, "new97-09.htm", h, "97.0년 9.0월"
  (R)19251.0, "new97-08.htm", h, "97.0년 8.0월"
  (R)44421.0, "new97-07.htm", h, "97.0년 7.0월"
  (R)17234.0, "new96-8.htm", h, "96.0년 8.0월"
(HL)/nal/whatsnew/
  (R)1774.0, "nal/2/lib2.htm", h, "도서관 이용안내"
(HL)/nal/2/
  (R)1517.0, "lib2-1.htm", h, "이용대상 및 시간"
  (R)1960.0, "lib2-2.htm", h, "소장자료"
  (R)7573.0, "lib2-3.htm", h, "열람실"
(HL)/nal/2/
  (R)2183.0, "lib2-4.htm", h, "참고봉사"
  .....
(HL)/~nalib/
  (R)0.0, "http://www.kordic.re.kr/", h, "연구개발정보센터"
  (R)0.0, "http://www.kordic.re.kr/~giis/kristal2/index.html", h, "KRISTAL-II"
  (R)0.0, "mailto:w3@nanet.go.kr", o, "w3 nanet.go.kr"
(HL)/~nalib/
  (R)0.0, "mailto:w3@www.nanet.go.kr", o, "w3 www.nanet.go.kr"
(HL)/
(HDIR)
    
```

**3.2 시각화 부분**

본 논문에서 개발한 웹 정보 시각화 도구인 WIV의 시각화 부분은 사용자에게 시각적으로 웹 사이트에 대한 정보를 보여주는 기능을 수행한다. 파서 부분에서 자동으로 생성된 계층적 구조 파일(HDIR)을 사용하여 웹 사이트에 대한 여러 가지 정보들을 시각적으



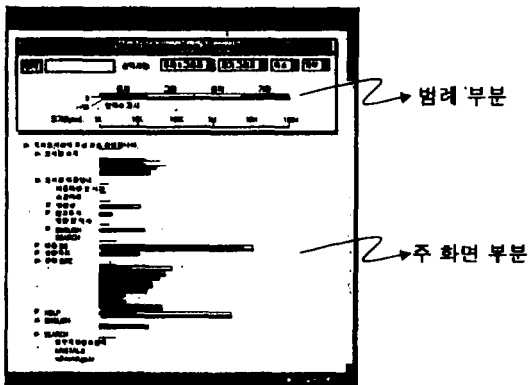
(그림 5) 위치 정보를 묻는 대화 상자  
(Fig. 5) Dialog box for asking location information



로 보여준다. 이렇게 정보를 시각적으로 보여주므로 파일이나 페이지들이 어디에 위치한지를 빨리 보여 주고 디렉토리들의 상대적인 크기들을 비교하고 장소에 잘 찾지 않는 계층적 구조의 큰 가치를 보기 위해 웹 사이트와 디렉토리를 브라우징하는 것을 도와 준다. 이러한 기능을 수행하기 위해서 시각화 부분은 웹 브라우저 상에서 실행되는 자바 애플릿(applet)으로 구현되어진다. 그러므로 자바를 지원하는 모든 브라우저에서 실행될 수가 있다. 시각화 부분은 크게 범례(legend) 부분과 주 화면(main) 부분으로 나눌 수가 있다. (그림 6)은 시각화 부분의 전체적인 모습을 보여주고 있다.

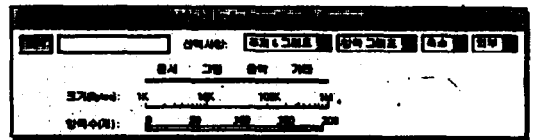
(1) 범례(legend) 부분

시각화 부분에서 범례 부분은 웹 정보에 대한 범례(legend)를 보여준다. 사용자가 선택적으로 정보를 볼 수 있도록 다양한 선택 사양을 위한 리스트 박스와 찾고자 하는 단어나 구를 검색하는 필드, 크기나 항목에 대한 척도를 제공해 줌으로서 시각적으로 상대적인 크기와 개수를 비교할 수가 있도록 하였다. 길이 막대 그래프에서 사용되어지는 색(color)은 파일의 종류를 표현해준다. 파일의 종류(문서, 그림, 음악, 기타)에 따라 색을 다르게 하므로 그 웹 페이지에서의 문서들의 종류가 어떻게 분포되어 있고, 그 상대적인 크기를 비교할 수 있다. 범례 부분에서는 이런 파일의 종류를 여러 색으로 구분하여 사용자에게 알려주고 있다.

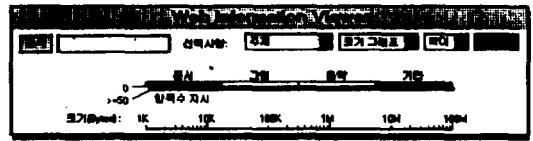


(그림 6) 시각화 부분의 전체적인 모습  
(Fig. 6) The overall display of visualization part

사용자의 선택 사항에 따라서 범례 부분은 약간 다르게 표현된다. 사용자가 선택 사항에서 항목의 개수를 시각적으로 표현한 항목 그래프를 선택하였을 때에는 (그림 7)과 같은 범례 부분이 생성된다. 크기 척도뿐만 아니라 항목 수에 대한 척도도 표시가 된다. 반면에 파일의 크기를 시각적으로 표현한 크기 그래프를 선택하였을 때에는 (그림 8)과 같은 범례 부분이 생성된다. 항목 수에 대한 척도는 사라지고 크기에 대한 척도만이 제공되며 항목의 개수는 길이 막대 그래프의 그림자로 표현된다.



(그림 7) 항목 그래프에 대한 범례 부분  
(Fig. 7) The legend part for a item graph



(그림 8) 크기 그래프에 대한 범례 부분  
(Fig. 8) The legend part for a size graph

범례 부분에서 선택할 수 있는 선택 사항은 다음과 같다.

- 주제/그래프/주제 및 그래프
- 항목 그래프/크기 그래프
- 확대/축소
- 내부/외부

또한 범례 부분은 마우스로 끌어서 이동이 가능하다. 주 화면 부분에서 항목의 내용이나 그래프의 크기가 분명치 않게 되는 것을 막기 위해서 전체적인 모양을 동적으로 수정할 수 있도록 하였다.

(2) 주 화면(Main) 부분

주 화면 부분은 계층적 구조 파일(HDIR)을 읽어 들여서 웹 사이트의 실질적인 개요 측면차를 보여주

는 부분이다. (그림 19)는 범례 부분과 주 화면 부분을 보여주고 각 부분에 대한 설명을 기술하고 있다.

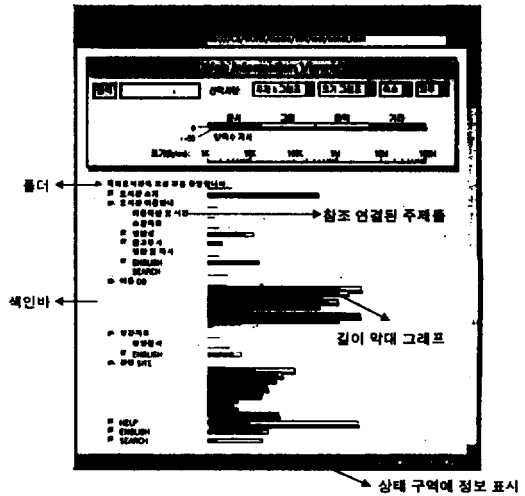
먼저 개요를 표시하기 위해서 참조 연결된 주제들을 계층적으로 표현하였다. 그러한 계층성을 시각적으로 보여주기 위해서 폴더(folder)와 색인바(index bar)가 사용되어진다. 확대된 레벨은 폴더가 열린 모양으로 색인바에 “-”가 표시되며 축소된 레벨은 폴더가 닫힌 모양으로 색인바에 “+”가 표시된다. 확대/축소 기능이 제공되어진다. 사용자가 닫힌 폴더를 마우스로 클릭하거나 색인바의 “+” 표시를 클릭하거나 길이 막대 그래프를 클릭하면 그 레벨에서 하위로 확대된다. 그러한 레벨이 중첩으로 존재할 경우, 같은 방법으로 확대할 수 있다. 또한 열린 폴더를 마우스로 클릭하거나 색인바의 “-” 표시를 클릭하면 상위 레벨로 축소가 된다. 이러한 기능은 사용자가 화면을 효율적으로 사용할 수 있도록 도와주며 같은 레벨을 묶어주어 편리한 검색 기능을 제공해 주는 클러스터링 전략에 적용된다.

각 주제의 오른쪽 길이 막대 그래프는 그 레벨 또는 페이지의 파일의 크기와 종류에 대해서 보여준다. 길이 막대 그래프는 범례 부분의 선택 사항에서 선택한 그래프로 표시된다. 즉 크기 그래프와 항목 그래프로 표시될 수 있다. 확대된 레벨에서의 그래프는 각 페이지의 파일의 크기와 종류를 보여주고 축소된 레벨에서의 그래프는 그 레벨 이하의 모든 파일의 크기와 종류 및 항목의 개수를 보여준다. 크기 그래프인 경우, 사용되어지는 색(color)은 파일의 종류를 표현해준다. 파일의 종류(문서, 그림, 음악, 기타)에 따라 색을 다르게 하므로 그 웹 페이지에서의 문서들의 종류가 어떻게 분포되어 있고, 그 상대적인 크기를 비교할 수 있다. (그림 10)은 항목의 개수를 나타내주는 항목 그래프에 대해서 보여주고 있다.

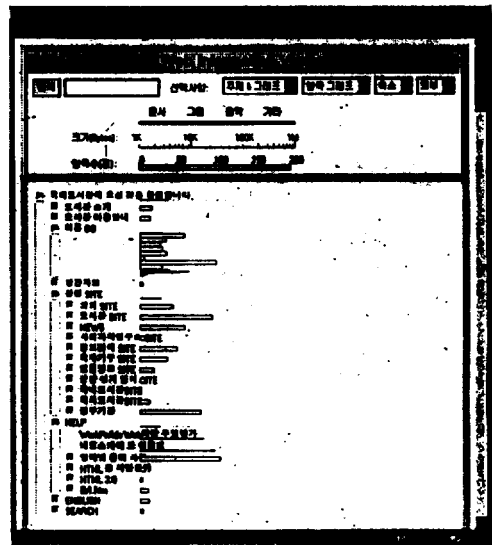
각 주제와 그 그래프는 일치하는 웹 페이지에 연결되고 계층적인 구조를 빠르게 검색하여 존재하는 페이지를 중간 단계의 페이지를 거치는 일 없이 빠르게 접근하게 해준다. 즉 기존의 브라우저와는 달리 개요만을 살펴보고 한 번에 원하는 페이지로 이동이 가능하다.

또한 문서의 수와 표현된 크기는 브라우저의 상태 구역(status area)에 사용자의 커서가 길이 막대 그래프 상에 있을 때 표시된다. 그러므로 시각적으로 뿐

아니라 정확한 수치상으로도 정보를 확인할 수가 있다.



(그림 9) 주 화면 부분에 대한 설명  
(Fig. 9) Description for a main display part



(그림 10) 항목 그래프  
(Fig. 10) Item graph

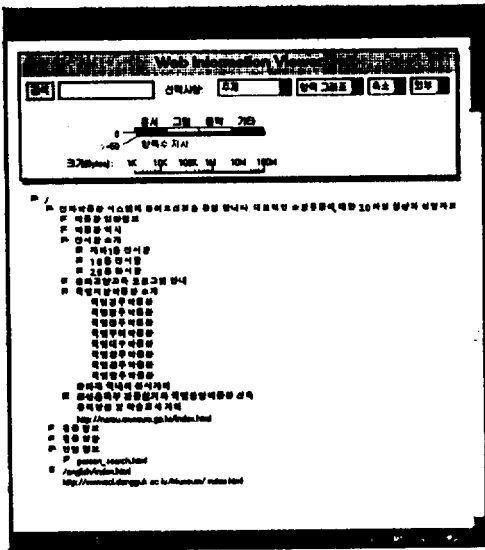
### 3.3 활용 방안

본 논문에서 개발한 웹 정보 시각화 도구 WIV는 다양한 방법으로 활용할 수 있다. 목차 기능, 분포 상

황을 파악하는 기능, 검색 기능, 웹 사이트의 개요 등 많은 부면에서 이용할 수 있다. 본 절에서는 계층적 구조 파일을 이용하여 시각화한 웹 정보들을 활용하는 예에 대해서 기술할 것이다.

(1) 목차 기능

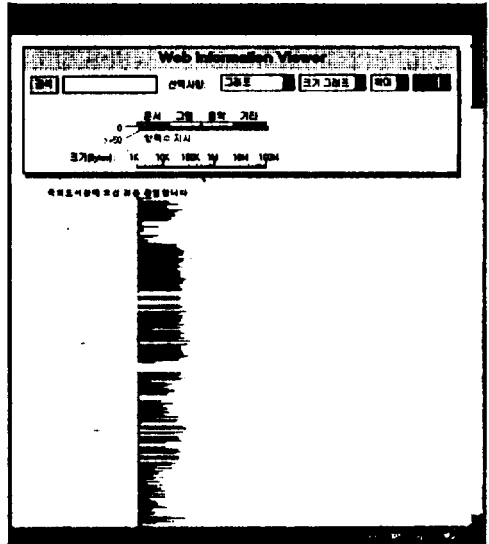
원하는 웹 사이트에 대한 정보들에 대해서 그래프 없이 주제로만 계층적으로 표현한다면, 보통의 책의 목차(table of content)와 같은 기능을 수행할 수 있다. 웹 사이트에 대한 정보를 책의 목차처럼 표현해주고 확대/축소 기능으로 편리하게 문서 공간을 검색할 수 있다. 참조 연결된 주제들의 효과적인 정보 제공을 위해서 관리자는 필터링을 할 수 있다. 문서 내용의 효율적인 체계적인 정보 제공을 위해 필요한 곳과 참조 연결되어 목차에 표시되기를 원치 않은 부분에는 <HL> 태그를 첨가하므로 웹 페이지를 파싱할 때, 필터링 되어질 수 있다. (그림 11)은 예로서 국립 중앙 박물관(http://www.museum.go.kr)의 사이트에서 목차 기능을 수행하는 것을 보여주고 있다.



(그림 11) 목차 기능  
(Fig. 11) Contents skill

(2) 분포 상황을 파악하는 기능

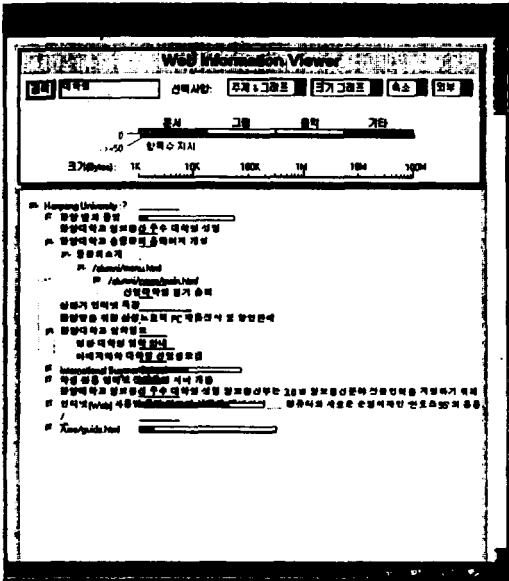
주제 없이 그래프로만 모든 세그먼트를 표시할 수 있다. 확대/축소 기능을 사용하여 그래프로만 표현하는 방법은 사용자가 웹 사이트의 시각적인 개요를 얻고 파일을 크기를 직접 비교하는 능력을 제공해준다. 또한 전체적으로 정보의 분포 상황을 파악할 수 있는 능력을 제공해준다. (그림 12)는 예로서 국회 도서관(http://www.nanet.go.kr)의 분포 상황을 파악하는 모습을 보여주고 있다.



(그림 12) 분포 상황 파악 기능  
(Fig. 12) Skill for understanding distribution

(3) 검색 기능

사용자가 찾고자 하는 단어나 구를 검색할 수 있는 기능을 제공한다. 큰문서 공간을 브라우징할 때 사용자는 자신이 원하는 정보를 얻기 위해 그 정보에 대한 특정 단어나 구를 입력하면, WIV는 그 단어나 구와 일치하는 모든 부분을 보여준다. 그러므로 사용자는 쉽게 자신이 원하는 정보를 중간 페이지를 거치지 않고 한 번에 이동하여 얻을 수 있다. (그림 13)은 한 예로서 한양대학교 웹 사이트(http://www.hanyang.ac.kr)에서 "대학원"에 대한 정보를 얻기 위해 검색하는 모습을 보여주고 있다.



(그림 13) 검색 기능(한양대학교)  
(Fig. 13) Search skill (Hanyang University)

용자가 방향 상실 없이 쉽게 검색할 수 있도록 할 수 있다. 본 WIV의 시각화 부분은 자바 애플릿으로 구현되어졌기 때문에 플랫폼에 관계없이 자바를 지원하는 모든 브라우저에서 사용할 수가 있다.

(그림 14)는 이러한 활용 예를 보여주고 있다. 국립중앙 박물관의 웹 사이트(<http://www.museum.go.kr>)를 파싱하여 얻은 계층적 구조 파일을 관리자가 프레임을 이용하여 웹 사이트의 개요로 추가함으로써 방문하는 사용자는 문서 공간의 분포 상황과 크기 비교 또한 손쉽게 자신이 원하는 정보를 얻을 수 있다.

4. 결론 및 향후 과제

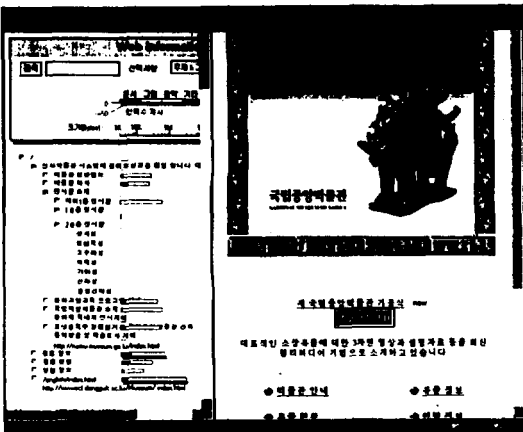
하이퍼텍스트 검색 작업은 방향 상실의 문제점을 가지고 있다. 이리므로 웹에서의 정보를 시각화하는 일이 필요한 일인데, 서로 연결된 문서들의 집합에서 정보의 양과 분포를 표현하는 일은 어려운 문제이다. 문서의 형태의 따라 시각화한 정보들(문서, 그림, 음악, 기타)은 사용자가 깊고 복잡한 구조를 브라우징 하는데 많은 시간을 소비하는 일 없게 해줄 수 있다.

본 논문에서는 이러한 문제점을 해결하기 위한 계층적 구조를 이용한 웹 정보 시각화 도구인 WIV (Web Information Viewer)를 개발하였다. 기존의 도구들의 문제점인 플랫폼 비호환성과 시각화 정보의 재사용의 어려움과 같은 문제들을 해결하였고 장점들을 수용한 네 가지 전략 즉 바인딩, 클러스터링, 필터링, 계층화를 통해서 이러한 문제점을 해결하고 웹에 대한 정보를 시각화할 수 있었다. 자바(Java)로 구현되어졌기 때문에, 하드웨어 독립적으로 사용되어질 수 있고 자바를 지원하는 모든 브라우저에서 사용할 수 있도록 설계되었다.

본 논문에서 개발한 웹 정보 시각화 도구인 WIV는 파일이나 페이지들이 어디에 위치한지를 빨리 보여주고 디렉토리들의 상대적인 크기들을 비교하고 평소에 잘 찾지 않는 계층적 구조의 큰 가치를 보기 위해 웹 사이트와 디렉토리를 브라우징하는 것을 도와 줄 것이다.

본 논문의 향후 연구 과제로써 다음을 고려해 볼 수 있다.

첫째, 본 논문에서 제시한 웹 정보 시각화 도구인 WIV는 웹 사이트에 대한 여러 가지 정보를 시각화하



(그림 14) 웹 사이트의 개요(국립중앙 박물관)  
(Fig. 14) The overview of web site(National Museum of Korea)

(4) 웹 사이트의 개요

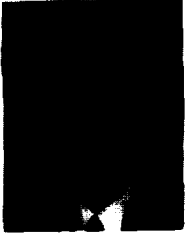
WIV의 파서 부분에서 원하는 웹 사이트를 파싱하여 자동적으로 생성해준 계층적 구조 파일(HDIR)을 이용할 수 있을 것이다. 웹 사이트의 관리자의 입장에서 이 계층적 구조 파일을 이용한 시각화 부분을 개요 형식으로 웹 사이트에 첨가함으로써 방문한 사

였다. 문서의 크기와 존재하는 항목 수 등을 막대 그래프를 사용하여 시각화함으로써 파일이나 디렉토리의 상대적인 비교와 분포상황을 알 수 있도록 하였다. 이러한 정보뿐만 아니라 웹 사이트에서 데이터를 갱신한 것을 알려주기 위해 갱신한 날짜와 시간을 표시해 주는 등의 좀 더 다양한 특성들을 시각화 할 수 있도록 설계할 수 있다.

둘째, 본 논문에서 개발한 도구에서 파서 부분을 자바 응용 프로그램이 아닌 자바 애플릿으로 작성하므로 자바 인터프리터 없이도 자바를 지원하는 모든 웹 브라우저에서 실행되도록 할 수 있다. 이렇게 되려면, 일단 자바 애플릿의 보안 측면이 개선되어야 할 것이다.

### 참 고 문 헌

- [1] Andrews, Keith, "Browsing, Building, and Beholding Cyberspace, New Approaches to Navigation, Construction, and Visualization of Hypermedia on the Internet," Doctoral dissertation, Graz University of Technology, September. 1996.
- [2] Asahi, Toshiyuki, Turo, David, Shneiderman, Ben, "Using Treemaps to Visualize the Analytic Hierarchy Process," Information Systems Research Vol. 6, No. 4, pp. 357-375, December. 1995.
- [3] 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.
- [4] Chimera, Richard, "Value bars: An Information Visualization and Navigation Tool for Multi-tribute Listings," Proc. CHI'92 Conference: Human Factors in Computing Systems, ACM, New York, pp. 293-294, 1995.
- [5] Chimera, Richard, Shneiderman, Ben, "An Exploratory Evaluation of Three Interfaces for Browsing Large Hierarchical Tables of Contents," ACM Transactions on Information Systems, Vol. 12, No. 4, pp. 383-406, October. 1994.
- [6] Egan, Dennis E., Remde, Joel R., Gomez, Louis M., Landauer, Thomas K., Eberhardt, Jennifer., Lochbaum, Carol C., "Formative Design-Evaluation of SuperBook," ACM Transactions on Information Systems, Vol. 7, No. 1, pp. 30-57, January. 1989.
- [7] Johnson, B., "Treemaps: Visualizing Hierarchical and Categorical Data," Doctoral dissertation, UMI-94-25057, August. 1993.
- [8] McKnight, C., Dillon, A., & Richardson, J., 'A Comparison of Linear and Hypertext Formats in Information Retrieval', In R. McAleese & C. Green(Eds.), Hypertext: The state of art(p.10-19). Oxford, England: Intellect Books, 1990.
- [9] Mukherjea, S., Foley, J. D., "Visualizing the World-Wide Web with the Navigational View Finder," Computer Networks and ISDN Systems Vol. 27, No. 1, pp. 1075-1087, 1995.
- [10] Rouet, J., Lovonen J., 'Studying and Learning with Hypertext: Empirical Studies and their Implication', In Rouet Jean-Francois, Lovonen Jarmo J., Dillon, Andrew, Spiro, Rand J.(Eds.): Hypertext and Cognition. Lawrence Earlbaum, New Jersey, 1996.
- [11] Spring, Michael B., Morse, Emile and Heo, Misook, "Docuverse: Multi-level Navigation of a Document Space," Department of Information Science and Telecommunications, University of Pittsburgh, Pittsburgh, PA 15260 USA.
- [12] Nahum D. Gershon, Frank Kappe "Visualizing the World Wide Web: Putting the User in the Driver's Seat," 4th International World Wide Web Conference, December 1995.
- [13] Elliotte Rusty Harold, 'Java Network Programming', O'reilly, 1997.
- [14] Ken Arnold, James Gosling, 'The Java Programming Language', Addison Wesley, 1996.
- [15] Prashant Sridharan, 'Advanced Java Networking', Prentice Hall PTR, 1997.

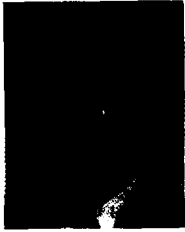


**노 원 빈**

1996년 한양대학교 전자계산학과 졸업(학사)

1996년~현재 한양대학교 전자계산학과 석사과정

관심분야: 네트워크 관리, 분산처리 시스템, 정보 시각화



**허 신**

1973년 서울대학교 전기공학과 졸업

1979년 Univ. of Southern California 전자계산학 석사 학위 취득

1986년 Univ. of South Florida 전자계산학 박사학위 취득

1986년~1988년 The Catholic University of America 조교수

1988년~현재 한양대학교 전자계산학과 부교수

관심분야: 분산처리 시스템, 결합하용 컴퓨터 시스템, 실시간 운영체제임