

# Java를 이용한 웹 검색결과 시각화 도구 개발

윤보열\*, 전형민, 정영아, 김응곤  
순천대학교 컴퓨터학과

## Development of a Visualization Tool for the Search Results on the WWW with Java

Bo-yul Yoon\*, Hyoung-min Joun, Young-a Jeong, Eung-kon Kim  
Dept. of Computer Science, Sunchon National University

### 요약

WWW상에서 많은 정보 가운데 자신에게 필요한 정보를 찾기 위해 검색 엔진을 사용하는데, 대부분의 검색도구들은 문자 중심의 스크롤되는 페이지로 보여주기 때문에 검색결과가 많아지면 원하는 정보를 쉽게 찾기 힘들고 앞 뒤 임의의 위치로 이동하기가 어렵다. 본 논문에서는 검색 결과를 시각화하여 보여줌으로써 편리한 사용자 인터페이스를 제공하여 원하는 정보를 쉽게 찾고, 그 문서가 있는 웹사이트로 곧바로 가는 기능을 가진 시각화 도구를 개발하였다. 본 시각화 도구는 Java로 구현하여 모든 플랫폼에서 사용할 수 있으며, Java를 지원하는 모든 웹브라우저 상에서 실행될 수 있다.

### 1. 서론

요즈음 인터넷을 통하여 많은 정보를 얻고 있다. 우리 주위에 정보가 폭발적으로 늘어나 홍수를 이루는데, 많은 정보 가운데 필요한 정보를 신속하고 정확하게 얻는 것이 무엇보다도 중요하다. 다행히 검색엔진이 개발되어 WWW상에서 원하는 정보를 쉽게 찾을 수 있도록 해주고 있다.

그러나 대부분의 검색엔진은 문자중심이며, 그 결과는 스크롤된 페이지로 나타나게 된다[1]. 문자는 그림보다 추상적이어서 많은 인지적 노력을 필요로 하며 사용자가 쉽게 접근할 수가 없다. 뿐만 아니라 결과가 길어지면 여러 페이지에 걸쳐서 순차적으로 출력되기 때문에 앞 뒤 임의의 페이지로 넘어가기가 어렵다[2,3].

따라서 본 논문에서는 새로운 시각화 도구를 만들어 사용자에게 편리한 인터페이스를 제공하고, 검색엔진으로부터 나온 검색결과를 화면에 시각화하여 보여줌으로써 원하는 정보를 쉽게 찾도록 하고자 한다.

WWW상에서 Client가 입력한 검색어는 Server로 들어가서 검색엔진을 통한 결과가 시각화 도구를 거쳐

HTML과 Java를 사용해 Client의 화면에 시각화되어 나타나게 된다. 본 시각화 도구는 알타비스타의 검색결과를 이용하고, Java를 사용하여 Client/Server 통신 환경을 구축하여 개발되었다.

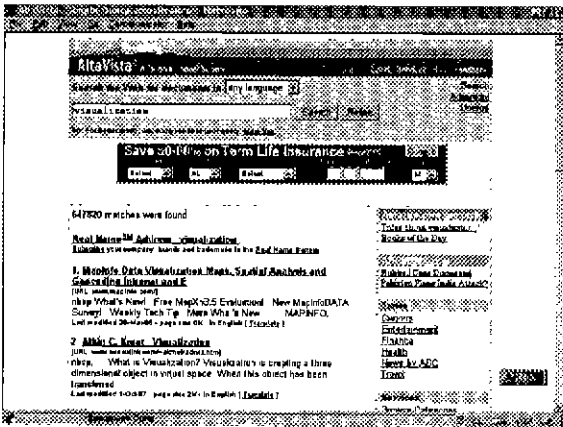
본 논문은 2장에서 기존 WWW 검색결과 출력에 있어서 문제점을 살펴보고 새로운 시각화 화면을 설계한다. 3장에서는 시각화 도구의 시스템 구조를 간략하게 살펴보고, 4장에서는 Java 네트워킹을 사용하여 WWW 상에서 구현하고, 5장에서 결론과 향후 연구 과제를 살펴보고자 한다.

### 2. 기존 검색결과 문제점과 본 연구의 시각화 도구

오늘날 대부분의 검색엔진은 문자 중심이어서 결과를 출력하는데 몇 가지 문제점들이 있다[3]. 첫째, 문자는 상징적이고 추상화된 개념이기 때문에 사용자가 육안으로 검색결과를 쉽고 자연스럽게 관찰할 수가 없다. 둘째, 스크롤된 문장들로 출력되므로 사용자에게 그것들을 통합해야 하는 인지적인 부담을 갖게 한다. 셋째,

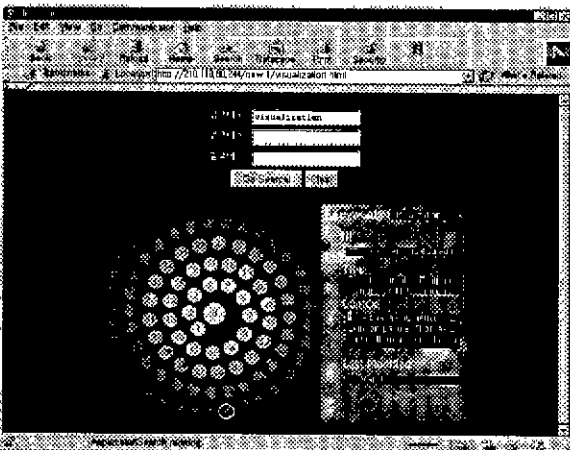
사용자는 전체 검색결과를 보지 못하고 화면에 출력된 부분적인 내용만 볼 수 있기 때문에 관심부분을 쉽게 찾거나 임의의 곳으로 즉시 갈 수 없다.

아래 <그림 1>은 알타비스타의 검색결과로 위에서 언급한 문제점들을 드러내고 있다. 문자열이 나열되어서 시각적으로 한 눈에 들어오지 않고, 10개씩 부분적으로 출력되면서 스크롤 되어버리기 때문에 전체를 보거나 다시 임의의 항목으로 가는 것이 어렵다는 것을 알 수 있다.



<그림 1> 알타비스타의 검색 결과

<그림 2>는 이런 몇 가지 문제점을 개선한 본 연구의 시각화 도구를 통한 검색결과를 보여주고 있다. 문자의 틀을 벗어나 도형으로 보여주므로, 사용자에게 시각적 부담을 덜어주고 색상이나 명암 정보를 통해 검색결과를 분류할 수 있으며 더 많은 정보를 화면에 출력할 수 있다.



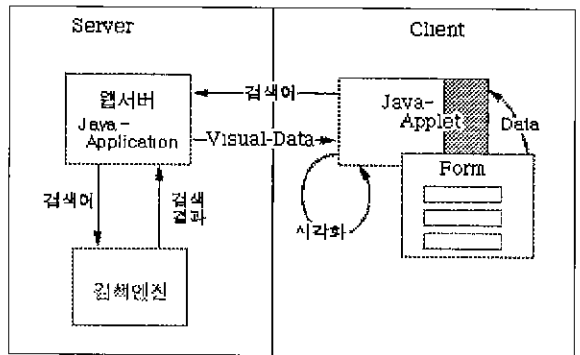
<그림 2> 시각화 도구를 사용한 검색 결과

입력 창에 검색어를 입력하면 검색결과항목은 자료의 우선 순위에 따라 나선형으로 번호가 붙으며 100개 출력된다. 지금까지 문장을 스크롤시키는 기존의 방법은 화면의 크기에 따라 디스플레이하는 항목의 수가 제약을 받았으나 여기서는 조그마한 아이콘을 써서 더 많은 것을 한 화면에 보여주고 있다.

마우스의 포인터가 각 항목위로 지나가면 그 항목에 대한 URL, 제목, 내용, 날짜, 크기 등 자세한 내용이 출력된다. 이때 마우스를 클릭하면 바로 그 URL의 사이트로 갈 수 있다.

### 3. 시각화 도구의 구조

사용자는 WWW상에서 검색어를 입력하게 되면 검색 결과는 시각화 도구를 통해 앞의 <그림2>처럼 출력된다. <그림3>은 입력한 검색어가 시각화 도구를 거쳐서 사용자의 화면에 영상으로 나타나는 과정을 보여주고 있다[4,5].



<그림 3> 시각화 도구의 구조

Client는 질의어를 입력받아 Java Applet에서 검색식을 만들어 Server의 Java Application을 통해 검색엔진으로 보내져 그 결과를 받아온다. 여기서 검색결과를 파싱하고 분석하여 시각화 데이터로 바꿔주며 Client에서는 HTML과 Java를 이용한 시각화 도구를 통해 원하는 결과를 시각화된 화면으로 볼 수 있게 된다.

### 4. Java 네트워크를 이용한 시각화 도구의 개발

Java를 이용하여 2장에서 설계한 검색결과항목의 시각화 화면을 웹브라우저를 통하여 볼 수 있도록 도구를 개발하였다. Java는 CGI와 달리 Server에 부담을 주지 않고, 어느 플랫폼이든 같은 결과를 보여준다. Java Applet은 인터넷 웹브라우저 상에서 실행되는 Java프로그램이다. 따라서 Applet은 상당한 보안 문제를 안고 있다[6].

Java Application으로 임의의 URL에 접속하여 데이터를 받아오는 것은 간단하지만 이것을 웹으로 옮기는 데에는 쉽지 않는 문제가 발생한다.

Applet은 자신이 실행된 Server 이외의 네트워크로는 연결이 불가능하다. 이는 검색엔진 같은 네트워크와 밀접한 관계가 있는 프로그램에서는 큰 문제가 아닐 수 없다. 다른 방법으로 검색 부분을 Application으로, 검색결과를 표시를 Applet으로 작성하여 Applet에서 Application을 불러 사용할 수 있으나 이것 역시 보안 문제가 생긴다.

결국 Client/Server 구조를 만들어 구현할 수 있었다. Client는 브라우저와 밀접한 관계가 있으므로 Applet으로 작성하고, Server는 네트워크가 가능한 Application으로 개발하였다.

특정 포트를 사용하는 Socket을 제작하고, 그에 따라 Client와 Server를 제작한다. 그리고 멀티 유저 환경을 위하여 서버 프로그램에서 검색 루틴을 따로 분리시킨다.

Server는 항상 Client의 접속을 다음과 같이 기다린다.

```
ServerSocket ss=newServerSocket(2501)
.....
Socket soc=ss.accept();
```

Client의 접속요청이 들어오면 Server는 soc라는 Socket을 생성하여 Client와 통신할 수 있도록 해준다. Client는 다음과 같이 Server에 접속요청을 한다.

```
Socket clientSocket = new Socket
(getCodeBase().getHostName(), 2501);
```

Client는 Applet으로 작성된 코드이므로 Server에서의 주소는 허용하지 않는다. Server를 어디에 설치하더라도 작동할 수 있도록 getCodeBase()를 사용하여 Server의 주소를 알아낸다.

Server와 Client간의 통신루틴인 Socket이라는 다리를 만든다. 이제는 이 다리를 건널 교통수단을 만들어야 한다. Socket은 자신을 입·출력 매개체로 삼을 수 있는 방법을 제공해 준다. 다음과 같이 입·출력 스트림으로 자료를 보냄으로써 Client/Server는 대화를 할 수 있게 된다.

```
InputStream is = new InputStream
(Socket.getInputStream());
OutputStream os = new OutputStream
(Socket.getOutputStream());
```

검색된 결과는 이 입·출력 스트림을 통하여 Server에서 Client로 보내어지고, Client는 이 자료를 바탕으로 시각화하여 화면에 보여준다.

## 5. 결론 및 향후 과제

검색결과를 영상으로 시각화하여 보여주는 도구를 설계하고 구현하였다. 이 시각화 도구는 사용자에게 편리한 인터페이스를 제공하고, WWW상에서 관심 있는 정보를 쉽게 찾는데 매우 유용하다.

검색결과를 스크롤되는 문장 대신에 시각화된 아이콘이 나선형으로 펼쳐져 원하는 것을 쉽고 빠르게 찾으며, 선택했을 때 그 문서가 있는 웹사이트로 즉시 갈 수 있다.

CGI대신 Java를 사용하였기 때문에 웹서버에 대한 부담을 줄이고, 어느 플랫폼에서든 사용할 수 있으며, Java를 지원하는 모든 웹브라우저 상에서 실행된다.

앞으로 검색결과를 3D로 구성하여 주어진 해상도를 가지고도 더 많은 정보를 시각화하며, VRML을 이용하여 더욱 생생하게 가상현실 공간을 돌아다니며 검색할 수 있도록 되어야 할 것이다[7]. 그리고 문자 뿐 만 아니라 이미지도 검색할 수 있는 데이터베이스 구축과 검색엔진을 개발하여 시각화 도구와 연동시킬 때 큰 효과가 있을 것이다.

## 참고문헌

- [1] Sougata Mukherjea, Kyoji Hirataand Yoshinori Hara, "Towards a Multimedia World-Wide Web Information Retrieval Engine", Sixth International WorldWide Web Conference, 1997
- [2] Marc H. Brown, Hannes Marais, Marc A. Najork, William E. Wehl, "Focus+Context Display of Web Pages", <http://gatekeeper.dec.com/pub/DEC...technical-notes/SRC-1997-010-html>, 1997
- [3] Lous C. Vroomen, "Information Visualization : An Overview", <http://www.crim.ca/~vroomen/writing/technical/reports/infovis.html>, April 1998
- [4] Jason Wood, Ken Brodlie and Helen Wright, "Visualization Over The World Wide Web And Its Application To Environment Data", IEEE, p81-86, 1996
- [5] Bo-yul Yoon and Eung-kon Kim, "The Design and Implementation of a Visualization Tool for the Search Results on the World-Wide Web", ICEIC '98, 1998
- [6] Elliotte Rusty Harold, "Java network programming", O'REILLY, 1997
- [7] Randall M. Rohrer and Edward Swing, "Web-Based Information Visualization", IEEE Computer Graphics and Applications, p.52-59, August 1997