웹 사용 데이터를 이용한 사이트맵 시각화에 관한 연구

Web Sitemap Visualization Using Usage Data

이명로, 정지홍
국민대학교 테크노디자인대학원 디지털미디어디자인학과

Lee, Myoung-Lo, Jung, Ji-Hong
Dept. of Digital Media Design, Graduate School of Techno Design, Kookmin University

• Key words: Sitemap, Information Visualization, Web Log Analysis, Node-Link Diagram

1. 서 론

월드 와이드 웹은 웹 페이지와 사용자와의 매우 복잡한 정보로 구성되어 있다. 이러한 복잡한 환경에서 사용자는 길을 잃어버리지 않기 위해 사이트맵을 이용한다. 사이트 맵은 현재 위치와 가야할 곳의 방향을 알려준다. 그러나 웹사이트의 정보가 점차 증가함에 따라 웹 페이지도 많아지게 되었다. 이에 따라 기존의 시각화 방법으로는 방대한 양의 자료를 사이트맵에 보여줄 수 없게 되었다. 이를 해결하기 위한 방법으로 사용자의 웹 사용 데이터를 이용한 3차원 노드-링크 다이어그램 사이트맵을 제안하고자 한다. 본 연구의 목적은 사용자가 웹 사이트를 방문하여 생성되는 양적 정보와 동적 정보의 시각화 방법을 제시하는데 있다.

2. 사이트맵의 이해

사이트맵은 웹마스터나 컨텐츠 제공자가 복잡한 웹 사이트를 한눈에 볼 수 있게 하고 세부적인 내용을 검색하는데 도움을 주기 위해 만들어 놓은 것이다. 이러한 사이트맵을 이해하기 위해 첫째로 만드는 주체를 기준으로 유형분석을 하고, 둘째로 시각화의 기본적인 모델링 기법인 노드-링크 다이어그램을 살펴보고, 셋째로 웹 사이트에서 사용 데이터를 추출하기 위해 이용되는 웹 로그 파일 분석을 알아본다.

2-1. 사이트맵 유형

사이트맵은 만드는 방법에 따라 관리자나 디자이너가 시각적 전략수단을 이용해 만드는 것과 데이터를 기반으로 알고리즘을 이용해 만드는 것으로 나눌 수 있다. 전자의 사이트맵에 사용되는 사이트맵은 크게 수평, 수직적 관계로 구성되는 위계적인 리스트, 하위 목록을 팝업이나 드롭다운 메뉴 방식으로 순차적으로 보여주는 진행형 제시, 원 형태로 배열하는 원형, 제시되는 정보에 은유적인 관계를 사용하는 은유의 방식으로 나눌 수 있다.

후자의 사이트맵은 프로그램의 알고리즘을 통해서 컴퓨터가 실시간으로 만드는데, 주요하게 사용되는 시각화 기법에는 왜곡(Distortion), 지리학(Geography), 3차원 공간, 노드-링크(Node-link), 트리맵(Treemap), 다이어그램등이 있다. 왜곡은 유클리드 기하학이 아닌 지수적으로 증가하는 좌표의 공간에 사이트를 매핑한다. 이 사이트맵 중앙의 노드는 가장 커 보이고 가장자리고 갈수록 작게 왜곡되어 보인다. 왜곡맵의 장점은 작은 공간 안에 많은 수의 노드를 넣을 수 있어 전체를 한번에 볼 수 있게하는 장점을 가지고 있다. 프럼(Plumb) 디자인사의 시각적 유의 사전(Visual tresaurus)은 하나의 단어나 구, 그리고 그와 관련된 모든 단어나 구를 3차원 공간에서 보여준다. 단어들이 멀리 갈수록 그 단어들은 점점 더 작아지고 희미해진다. 처음 사용할때는 호기심을 유발하지만 너무 많은 클릭을 해야하기 때문에 오랫동안 사용할 수 없는 단점이 있다. 카르티아 씸스케이프(Cartia Themescape)가 제작한 정보 섬에서는, 왼쪽 창에 제어할 수 있는 세트가 있고, 각 항목을 클릭하면 이에 대응하는 정보가 지도 표면을 따라 검색되고 왼쪽 세트에서 선택한 항목이 지도 위에 점으로 나타난다.

2-2. 노드-링크 다이어그램

다이어그램(Diagram)은 단순한 등호나 기호를 사용하여 어떤 사물현상이나 상호관계를 도해하거나, 사물의 대체적인 형태와 여러 부분의 관계를 보여 주는 설명적인 그림이다. 정보화 사회에서 중요한 쟁점은 방대한 자료의 정보화 작업에 달렸는데, 다이어그램은 인간의 생활에서 일어나는 다양한 변화를 수치적, 수량적 정보로 전달하는 매체로서 중요한 위치를 점하고 있다.

노드-링크 다이어그램으로 가장 일반적인 데이터 모델은 엔티티-릴레이션 모델이다. 이 모델은 컴퓨터 공학과 비즈니스 모델링에서 넓게 사용된다. 엔티티-릴레이션에서, 엔티티 모델링에서 객체는 물체의 부분이 될수 있거나 혹은 조직에서의 한 부분과 같은 추상적인 것이 될 수 있다. 관계는 엔티티를 사이에 존재하는 다양한 종류의 연결이다. 예를 들어, 매장에서 고객과 판매자는 관계와 같다.

노드-링크 다이어그램은 2차원 맵위에 웹 사이트를 시각화 하기 위해 일반적으로 사용되는데, 각 페이지는 노드로, 페이지와 페이지 사이의 수평, 수직 관계는 링크로 사용된다.

2-3. 웹 로그 분석

웹 사이트를 운영하거나 자주 한 곳의 사이트를 가게되면 이 사이트에 누가 오는지, 어느 페이지를 많이 보는지, 언제 많이 오는지와 같은 현재의 사이트의 상황을 알고 싶어한다.

이러한 정보는 웹 서버에 로그파일로 기록된다. 각각의 상황에 따라 다른 파일로 기록되는데 사용자가 웹페이지를 접근할 때 기록되는 접근로그(Access log), 링크된 페이지가 없거나 CGI프로그램에서 에러가 날 때 작성되는 에러로그(Error log),

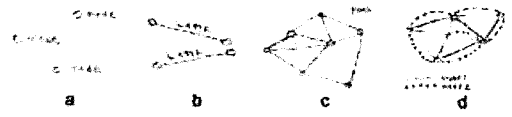
그리고 사용자가 간 경로을 기록하는 레퍼럴 로그(Referral log)등이 있다.

웹 로그 파일 분석을 통해 측정할 수 있는 내용은, 페이지에 연결된 파일의 숫자인 히트(Hits), 방문자가 요청한 페이지만 해당하는 페이지뷰(Page view), 특정 웹 페이지에 머무르는 시간인 체류시간(During time), 웹사이트에 접속해서 연속적으로 페이지를 본 후 다른 사이트로 이동하는 과정을 하나의 방문으로 기록하는 세션(Session), 특정 웹 사이트에서 한번 이상 접속한 사용자인 방문자(Visitors)이다.

3. 사이트맵 시각화 방법

3-1. 구조

노드[그림 3-1 a]는 사이트맵에서 가장 기본적인 요소이다. 하나의 노드는 각각의 페이지를 의미한다. 링크[그림 3-1 b]는 두 개의 노드사이의 연결이다. 보통 링크에 대한 인터랙션으로 두 노드가 강하면 근접해 있고 약하면 거리가 멀리 있게 각 노드의 힘을 조정한다. 길[그림 3-1 c]은 사용자가 걸어간 발자국으로 사용자가 걸어 다니는 길을 알 수 있어 매우 중요하다. 행위자[그림 3-1 d]는 같은 길을 여러 사람이 걸어갈 때 차이점을 보여 주게 한다.



[그림 3-1] 구조

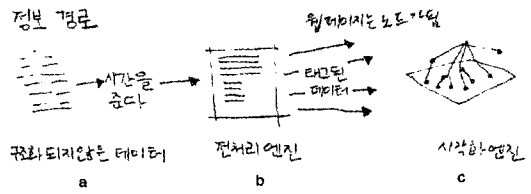
3-2. 시각화에 사용되는 값

시각화에 사용되는 값은 웹 로그 분석을 통해 얻어진다. 추출되는 값은 사용자, 방문횟수, 체류시간, 접속 시간, 경로이다. 체류시간 노드에, 경로는 링크에 사용된다.

3.3 시각화 프로세스

사이트맵의 구현을 위한 정보의 처리 과정은 아래와 같다.

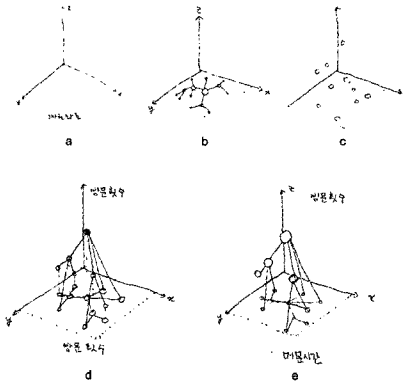
구조화 되지 않은 정보[그림 3-2 a]는 사용자의 웹 사이트 이용 데이터를 알 수 있는 웹 로그파일과 같이 특정한 목적으로 정보가 정리되지 않은 것이다. 전처리 엔진[그림 3-2 b]은 웹 로그파일과 같이 구조화되지 않은 정보를 노드, 링크, 경로, 사용자 등으로 구조화하는 것이다. 시각화 엔진[그림 3-2 c]은 사용자 이용 데이터를 시각화하기 위한 방법을 제시하고 시뮬레이션하는 과정이다.



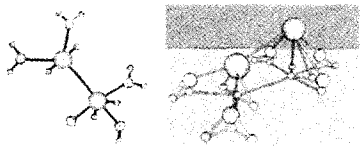
[그림 3-2] 시각화 프로세스

3.4 사이트맵 시각화 방법 제시

사용자의 이용 데이터는 노드-링크 다이어그램을 통해 3차원으로 매핑 된다[그림 3-3 a]. 시각화의 첫 번째 단계는 기본적인 사이트의 구조를 x-y평면에 매핑한다[그림 3-2 b]. 이것은 사이트가 계속 변하게 될 때 기준이 된다. 두 번째 단계는 사용자의 사용 데이터를 시각화 하는 단계로 방문횟수는 x-y평면에서 페이지에 해당하는 노드가 z축으로 이동하도록 시각화한다[그림 3-3 c]. 경로는 한 세션 동안에 이동 노드 사이를 연결하는 링크로 시각화 한다[그림 3-3 d]. 방문시간은 노드의 크기를 증가시킨다[그림 3-3 e]. 사용자는 크게 현재 사용자와 전체 사용자로 제한하고 컬러를 이용하여 구분한다. 또한 사용자가 최근에 간 흔적과 오래된 흔적에는 각 노드의 명도를 이용하여 오랜된 흔적은 흐리게 보이도록 한다. [그림 3-4]는 위의 방법을 이용하여 시각화한 사이트맵이다.



[그림 3-3] 사이트맵 시각화 방법



[그림 3-4] 사이트맵 시각화 예

4. 결론

웹 사용 데이터와 노드-링크 다이어그램을 기반으로 하는 시각화 방법은 사용자에게 전체적인 사이트의 구조를 개관적으로 빠르게 전달해 줄 수 있을 것이다. 또한 사용자에게는 사이트에서 일어나고 있는 상황을 알수 있게 할 것이다. 관리자에게는 사용자의 행위와 목적을 분석하고, 웹 사이트 사용성을 분석하고 예측하는 기본 자료로 사용될 수 있을 것이다.

참고문헌

- Colin Ware, Information Visualization, Morgan Kaufmann, 2000
- Maia Engeli, bits and spaces, Birkhauser, 2001
- *Benjamin Fry, Organic Information Design, M.S, 2000*
- 폴칸외 1명, 웹사이트 매핑, 안그라픽스, 2001
- http://www.cybergeography.org/atlas/atlas.htm 사이버지도
- http://www.imv.is.ocha.ac.jp:8080/~rika2/iv/웹 사이트 시각화기법
- *http://mappa.mundi.net* 가상지도