

웹 서버 성능 향상을 위한 링크 정보 분류의 링크 맵 시스템 구현

문일형*, 조동섭*
이화여자대학교 컴퓨터공학과*

Link Map System Implementation using link information classification for web server performance elevation

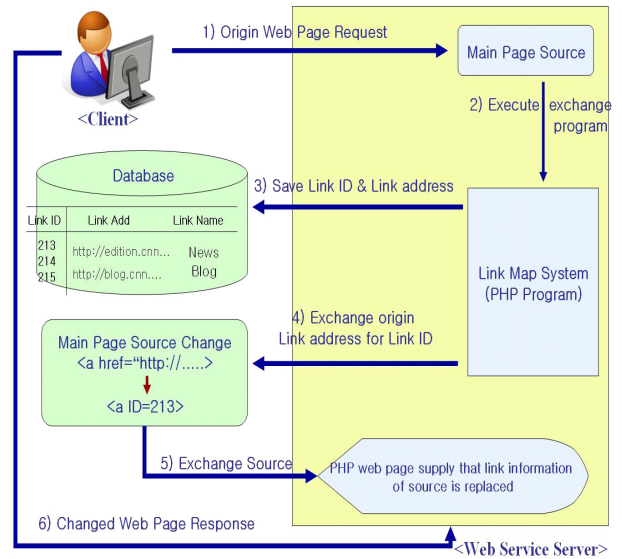
Yil-hyeong Mun, Dong-sub Cho
Dept. of Computer Science and Engineering, Ewha Womans University

Abstract - 포털 사이트와 더불어 여러 상업용 사이트의 경우 수많은 고객들을 수용하기 위해 여러 가지 서비스들을 개발하고 있다. 여러 가지의 서비스들은 사이트의 링크 서비스로 서비스 되고 있다. 그러나 이런 서비스의 개수들이 크게 늘어나면서 초기 서비스 웹페이지의 크기도 늘어나게 된다. 많은 링크 서비스들은 초기 웹페이지의 소스의 크기를 크게 함으로써 웹서버에 부하를 주게 된다. 다수의 사용자가 웹 페이지에 접근했을 경우 서비스 속도에 저하를 줄 수 있다. 이런 문제점을 보완하기 위해 링크 맵 시스템을 구현하여 링크 서비스 모두를 서비스 하는 것이 아니라 일부 서비스만 제공하는 것으로 서버의 부하를 줄일 수 있다. 접근 사용자들은 평소와 다를없이 웹 사이트에 접근하지만 링크 맵 시스템이 자동으로 불필요한 링크들을 제거하고 웹 페이지를 서비스한다. 이때 서버에 부하에 미치는 영향도 테스트하여 확인한다.

접근하게 된다. 이때 링크 맵 시스템이 구동되면서 웹 페이지의 소스를 변환하는데, ''태그를 '<a id>' 태그로 변환하여 하이퍼링크 기능을 비활성화로 변환한다. 또한 그에 연결된 링크 주소도 데이터 베이스에 저장되면서 데이터 베이스에서 생성된 ID로 대체하게 된다. 아래 그림은 링크 맵 시스템의 설계도이다.

1. 서 론

많은 인터넷 서비스 이용자들은 여러 종류의 웹 사이트에서 수많은 서비스를 제공 받는다. 이런 이용자들을 위하여 웹 서비스를 하는 웹 사이트에서는 여러 가지 종류의 서비스를 동시에 제공하고자 노력한다. 이런 이유로 초기 이용자들이 접근하는 웹페이지에 많은 서비스들이 하이퍼링크를 통해 제공된다. 그러나 이런 하이퍼링크 서비스의 증가는 웹 서버의 부하를 증가 시킬 수 있다. 그것은 웹페이지의 첫 페이지의 소스가 크게 증가하기 때문에 다수의 이용자들이 접근했을 때, 서버의 부하에 영향을 미치기 때문이다. 이런 점을 고려했을 때, 서버의 부하를 줄이기 위한 방법을 찾아야 한다. 본 논문에서는 이런 방법의 하나로 링크 맵 시스템을 구현하여 하이퍼링크를 제거하도록 하였다. 즉, 일반 사용자가 웹 사이트에 접근했을 때, 사용자가 주로 접근하는 링크만 남겨두고 나머지 링크는 제거하기 위한 링크 맵 시스템을 구현하였다. 이 시스템을 통해 하이퍼링크 서비스의 변환을 하게 되고, 초기 웹페이지의 소스의 크기가 줄어들게 된다. 그에 따른 웹서버의 부하도 감소하게 된다. 웹 서버의 부하는 더 많은 사용자들의 수용이 가능하게 되고 웹 페이지를 운영하는 사업자에게도 필요한 서비스가 될 것이다.



<그림 1> 링크 맵 시스템 구조

2. 본 론

2.1 관련연구

인터넷 서비스의 기하급수적 증가로 여러 종류의 사용자의 니즈에 맞는 웹 사이트들을 찾을 수 있게 되었다. 현재 많은 수많은 웹 페이지들 가운데 사용자들의 기호에 맞는 웹 페이지들의 관리를 위해 웹 페이지의 링크 정보들을 이용한 연구 및 프로그램들이 개발되었다. LinkManager 같은 프로그램은 다양한 종류의 웹 링크들을 구분하여 관리하도록 개발되어졌다.[1] 또한 대용량의 웹 문서의 효과적인 관리를 지원하는 관리서버의 연구로 이어진다. 이때 다량의 웹 문서들 가운데 좀 더 효과적인 관리를 위해 링크정보를 이용한 관리서버 개발이 연구되고 있다.[2] 이렇게 다양한 형태로 그 수가 기하급수적으로 많아진 링크들은 끊어진 링크, 즉 서비스를 더 이상 하지 않는 링크로 상대가 변하거나 사용하지 않는 파일 등으로 연결되어 그 기능을 상실한 링크들도 존재하게 되었다. 이런 링크들을 추출하여 새로운 링크로 수정하는 연구도 시작됐다.[3] 위와 같은 모든 연구들은 웹 페이지의 대부분을 링크들이 대부분 차지하게 됨으로써 생기는 문제들을 해결하기 위해 시작된 연구로서 앞으로 인터넷 웹 페이지들의 서비스 향상에 중요한 부분이 될 것이다.

2.2 설계 및 구현

본 절에서는 기존 웹 페이지의 소스의 링크 서비스 부분을 변환하는 기능을 구현하고자 한다. 링크 맵 시스템이 그 기능을 담당하는데, 기존의 웹페이지의 소스 중에 하이퍼링크부분의 정보를 태그와 주소들을 변환하는 작업을 하게 된다. 일반 사용자들은 평소와 같이 웹 페이지에

먼저 사용자가 웹 사이트에 일반적인 방법으로 접근 한다. 이때 메인 페이지는 기존의 웹 페이지 소스를 구동하는 것이 아니라 링크 맵 시스템에서 웹 페이지 소스를 가지고 변환 작업을 시작한다. 의 태그의 변환이 시작되는데, 의 태그가 로 변환하면서 뒤에 연결된 주소는 이미 생성된 데이터 베이스 테이블에 저장된다. 이때 Link ID와 Link Name이 동시에 생성된다. Link ID는 링크 주소들을 차례대로 번호를 주게 되는 ID가 된다. Link Name은 이 링크 서비스의 서비스 이름이 기재되어 있을 경우 저장된다. 그런 다음 Link ID는 다시 링크 맵 시스템에서 가공되어 로 웹페이지를 변환한다. 이렇게 변환된 웹페이지 소스가 웹 사용자에게 제공된다.

2.2.1 링크 맵 시스템 구현

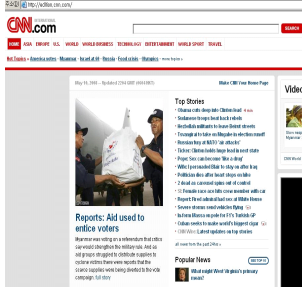
링크 맵 시스템은 PHP로 구현되었다.

<표 1> 링크 맵 시스템의 구현을 위한 환경 설정

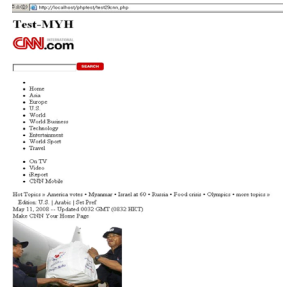
Web Server	Apache
Program Language	PHP
Database	MySQL
Test Website	http://edition.cnn.com/

위 표는 링크 맵 시스템 구현을 위한 환경 설정 부분이다. 링크 맵 시스템의 구현 테스트를 위한 사이트는 CNN 뉴스 사이트를 이용하였다. 이 사이트의 초기 웹페이지 소스를 이용하여 링크 맵 시스

맵을 구동한다.



〈그림 2〉 기존 CNN 초기 웹 페이지



〈그림 3〉 변환된 CNN 웹 페이지

위 그림은 링크 맵 시스템으로 변환된 CNN 웹페이지의 모습이다. 기존의 웹페이지의 다수의 링크들이 변환된 웹 페이지에서는 비활성화된 것을 쉽게 알 수 있다.

```




```

〈그림 4〉 기존 CNN 웹페이지의 소스

위 그림은 기존 CNN 웹페이지의 소스이다. 보다시피 많은 하이퍼링크들을 통해 서비스가 연결되어 있다. 이것은 웹페이지 소스의 대부분을 차지하게 된다. 이런 소스의 대부분을 차지하는 링크 서비스들은 초기에 모두 제공하기 보다는 사용자가 필요한 서비스만을 제공하고 그에 따라 서비스 되지 않는 링크들은 비활성화 하여 소스의 크기를 줄일 수 있다.

```

<div class="cnnHeader">
  <div class="cnnHeaderLeft">
    <a href="http://www.cnn.com/" class="cnnHomePage">Home</a>
    <a href="http://www.cnn.com/US/">US</a>
    <a href="http://www.cnn.com/EUROPE/">Europe</a>
    <a href="http://www.cnn.com/ASIA/">Asia</a>
    <a href="http://www.cnn.com/AFRICA/">Africa</a>
    <a href="http://www.cnn.com/TECH/">Technology</a>
    <a href="http://www.cnn.com/SPORT/">World Sport</a>
    <a href="http://www.cnn.com/FRANCE/">France</a>
  </div>
  <div class="cnnHeaderRight">
    <a href="http://www.cnn.com/US/">US</a>
    <a href="http://www.cnn.com/EUROPE/">Europe</a>
    <a href="http://www.cnn.com/ASIA/">Asia</a>
    <a href="http://www.cnn.com/AFRICA/">Africa</a>
    <a href="http://www.cnn.com/TECH/">Technology</a>
    <a href="http://www.cnn.com/SPORT/">World Sport</a>
    <a href="http://www.cnn.com/FRANCE/">France</a>
  </div>
</div>

```

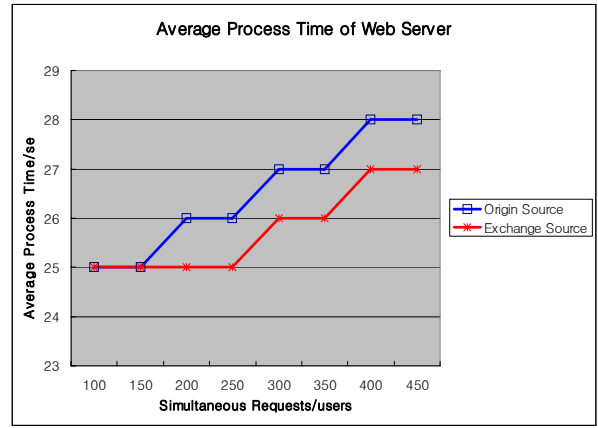
〈그림 5〉 변환된 CNN 웹페이지의 소스

링크 맵 시스템으로 변환된 웹페이지의 소스는 위 그림과 같다. 다소 길었던 링크 태그들은 위와 같이 변환되어 소스의 크기를 줄이는데 기여한다. 기존 CNN의 소스가 88.8Kb에서 변환 후, 83.7Kb로 줄어 들었다.

2.3 성능 평가

위와 같이 변환된 웹페이지 소스에 대한 서비스가 웹서버에 어떻게 영향을 미쳤는가에 대한 평가를 한다.

100개의 세션을 시작으로 10번의 동시 접속의 테스트를 2초 간격으로 웹서버에 했을 경우 Response Time을 측정하였다. 측정된 결과는 다음 그림과 같다.



〈그림 6〉 기존 웹페이지 소스와 변환된 웹페이지 소스에 대한 웹 서버 Response Time 측정 그래프

위 그림을 보면 기존의 웹페이지의 반응 시간보다 변환된 반응 시간이 더 빠르다는 것을 쉽게 알 수 있다. 이것은 웹서버에 미치는 영향을 바로 보여주는 것으로, 다수의 링크들에 관한 관리가 필요하다는 것을 보여준다.

3. 결론

링크 맵 시스템을 통해 현재 웹 서비스 되고 있는 페이지들에 대한 서비스 관리가 가능하다. 이는 많은 웹 서비스들이 링크들을 통해 이루어지고 있는 가운데, 웹페이지의 소스의 크기를 늘어나도록 하는 원인이 되기도 하다. 다시 말하자면, 웹 서버에 대한 부하도 증가 할 수 있음을 보여준다. 증가하고 있는 링크들을 링크 맵 시스템을 통해 관리하고 서비스함으로써 사용자들은 필요한 서비스만을 제공 받으면서 웹 서버 부하에 대한 서비스 저하를 피할 수 있으며, 운영 및 서비스 하는 웹 서버쪽에서도 웹 서버의 부하를 조금이나마 줄일 수 있다. 이는 더 많은 사용자들의 접근을 가능하게 하고, 필요한 링크들만을 서비스함으로써 고객의 니즈를 더욱 고려할 수 있게 된다.

더 많은 사용자들이 만족하는 서비스를 받기 위해서는 각각의 사용자들이 필요로 하는 서비스들의 구성을 어떻게 할 것인가의 문제가 대두 될 것이다. 즉 수많은 사용자들의 니즈를 어떻게 찾아내며 어떻게 서비스할 것인가에 대한 연구가 더 필요하다. 이런 서비스의 구성은 곧 웹서버의 부하에 미치는 영향과 무관하지 않기 때문에 심도 있게 연구되어야 할 것이다.

참고 문헌

- [1] 신원준, 정혜영, 박양수, 이명준, “웹기반의 효과적인 링크관리자”, 한국정보과학회 춘계학술대회, VOL.30, NO.01, 2003.04
- [2] 김영승, 박인석, 현순주, “가상 문서의 효율적인 관리를 지원하는 대용량 링크정보 관리서버의 개발”, 한국정보과학회 춘계학술대회, VOL.29 NO. 01 pp.175-177, 2002.04
- [3] 엄정섭, 유대승, 심민석, 이명재, “웹 페이지 관리를 위한 링크 추출과 검증”, 한국정보과학회, 추계학술대회, VOL.28 NO.02 pp.427-429, 2001.10
- [4] Yu Bo, Qi Luo. “Personalized Web Information Recommendation Algorithm Based on Support Vector machine”, International Conference on Intelligent Pervasive Computing, pp487-pp490, 2007
- [5] Lan Nie, Brian D. Davison, Xiaoguang Qi. “Topical Link Analysis for Web Search”, SIGIR04, August 6-11, pp91-pp98, 2006
- [6] J.M. Kleinberg. Authoritative sources in a hyperlinked environment. Journal of the ACM, 46(5):604-632,1999
- [7] 이수철, 이시은, 황인준, “웹 접근성 향상을 위한 웹 서핑 도우미”, 정보과학회논문지 소프트웨어 및 응용 제31권 제9호, 2004.9