

웹 서버 성능 향상을 위한 링크 맵 웹 시스템 구현

문일형*, 조동섭

이화여자대학교 컴퓨터공학과

llover78@gmail.com , dscho@ewha.ac.kr

Web Link Map implementation for Web Server performance improvement

Yilhyong Mun, Dongsub Cho

Ewha Womans' University Computer Science and Engineering

요 약

현재 수많은 상업포탈 사이트들은 다양한 서비스를 제공하기 위해 많은 노력을 기울이고 있다. 다양한 서비스를 제공함으로써 다양한 소비자들을 더 수용하기 위해 다양한 기능들을 개발하고 있다. 이러한 기능들을 제공하기 위해 많은 양의 하이퍼링크로 연결된다. 이런 많은 양의 하이퍼링크들은 초기 서비스 웹 페이지 소스의 대부분을 차지하게 된다. 이는 웹 서버의 부하를 주는 것은 물론이고 사용자들 역시 그렇게 많은 링크 서비스를 필요로 하지 않기 때문에 많은 양의 링크들을 정리할 필요가 생긴다. 즉 사용자들이 빈도수가 높은 서비스에 대한 링크만을 연결함으로써 웹 서버의 부하를 줄이고 웹 페이지의 소스의 양도 크게 줄일 수 있다. 일반 웹 페이지의 소스를 렌더링하여 필요한 링크만 제공하는 구현을 통해 웹 서버의 성능 향상을 연구한다.

1. 서 론

수많은 웹 사용자들은 수많은 웹 페이지의 서비스를 제공받고 있다. 그러나 너무나 많은 웹 페이지의 서비스들은 하이퍼링크로 연결되어 서비스 되고 있지만 모든 서비스를 모든 웹 사용자들이 이용하는 것은 아니다. 그러나 많은 웹 서비스 페이지들은 좀 더 많은 사용자들을 수용하기 위해 많은 기능의 서비스들을 링크를 통해서 제공하고 있다. 이런 경우 초기 웹 페이지의 서비스의 증가로 링크들의 정보는 더욱 더 많아지게 되고 그로 인해 초기 웹 페이지의 실행에 부하를 줄 수 있다. 즉 웹 서비스를 제공하는 웹 페이지의 사이즈가 증가하게 되고 서비스하는 속도라던가 사용자들의 이용이 적은 웹 링크들은 오히려 사용자들로 하여금 불편을 제공할 수 있다. 이런 문제점을 보완하고자 서비스 되고 있는 기존의 웹 페이지를 렌더링하고 수많은 링크들이 정리된 웹 페이지로 변환되는 구현이 필요하다.

본 논문에서는 링크 맵 서비스의 설계와 구현을 통해 웹 서버 향상에 대한 연구를 표현하고 분석한다.

2. 관련연구

인터넷의 사용은 기하급수적으로 늘어감에 따라 그에 따른 서비스도 늘어나게 되었다. 이렇게 늘어난 서비스들은 하이퍼링크를 통해 제공 되고, 그에 따라 웹

페이지의 구조는 복잡해지고 콘텐츠의 양도 늘어나게 되었다. 기존의 연구에서는 서비스 링크의 구조와 추출에 대한 연구를 통해 웹 검색의 성능을 향상하고자 하였다.

2.1 웹 접근성 향상을 위한 정보 검색 연구[3]

웹사이트에서 다루는 정보의 양의 증가로 인해 사용자가 사이트의 정보에 접근하기 위해 방문하는 첫 페이지에서 필요한 정보를 바로 수집하는 경우보다 여러 링크 경로를 거친 후에야 비로서 정보와 서비스에 접근하기 때문에 원하는 정보에 바로 접근하기 쉽지 않다. 또한 하이퍼링크를 이용하는 페이지가 많기 때문에 사용자가 해당링크에 연결된 페이지에서 다루는 정보에 대해 파악하는 것이 중요하지만, 링크 페이지의 콘텐츠를 미리 파악하는 것은 어렵고 부정확하다. 이런 문제점들은 실제 웹 저작환경과 웹 탐색도구들이 다루는 정보 단위의 차이로 인해 발생한다. 이런 웹 저작 경향을 반영하여 웹의 정보 접근을 어려움을 해결하기 위해 실제 정보의 단위를 정의 하고 이를 검색과 북마크, 브라우저의 단위로 활용하는 Link Broker시스템을 제안한다.

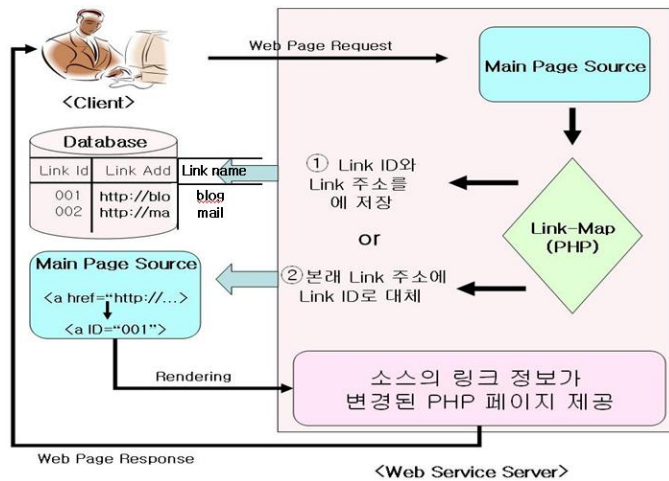
이 시스템은 논리적으로 연결된 정보를 담은 페이지들을 추출하고 그 뒤에 페이지셋으로 묶어 검색과 북마크의 단위로 만든다. 그리고 페이지셋에 속한 페이지 내의 주요 콘텐츠 구역들을 이용하여 컬렉션 페이지를 생성하여 흩어져있는 정보를 하나의

페이지에 표현한다. 다양한 검색 목적에 부합하는 페이지셋 단위의 검색과 북마크의 효율성의 향상과 컬렉션 페이지를 통한 정보 접근성의 향상이 가능하다.

3. 설계 및 구현

3.1 링크 맵 웹 시스템 설계

링크 맵 웹 시스템은 위 관련연구에서 서술된 Link Broker처럼 링크들을 여러 단위로 구분하여 표현하는 것이 아니라 수많은 사용자가 주로 사용하는 소수의 링크만을 기능하도록 하고 그 외의 많은 링크들은 하이퍼링크 정보를 삭제함으로써 초기 웹서비스 페이지의 사이즈를 줄이고, 웹 서버의 부하도 줄일 수 있다. 또한 초기 웹 페이지의 좀 더 빠르고 효율적인 로딩으로 더 많은 사용자들의 접근을 증가시킬 수 있다.



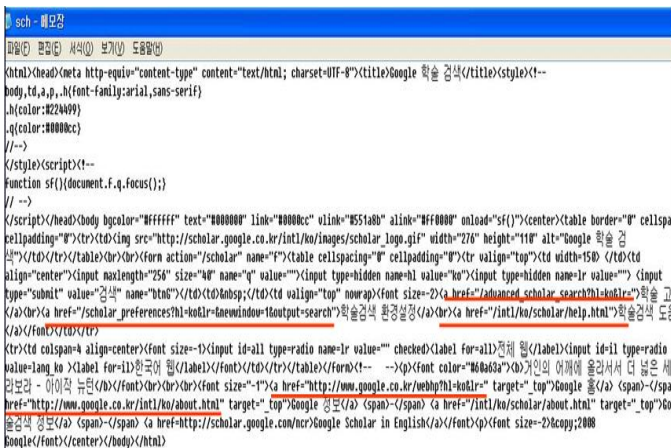
(그림 2) 링크 맵 PHP 프로그램

위 프로그램을 이용하여 웹 페이지가 렌더링 작업을 하게 된다. 테스트 페이지는 구글 학술검색 페이지의 소스를 링크 맵으로 렌더링 한다. 먼저 렌더링 전의 <http://scholar.google.com>의 웹 페이지를 링크 맵으로 렌더링하여 php파일로 변환하고 그 후 한번 더 사용자가 사용하지 않는 링크는 비활성화하도록 한 번 더 렌더링 한다.



(그림 3) <http://scholar.google.com> 의 메인 페이지

위의 메인 페이지의 소스를 링크 맵 PHP 프로그램으로 1 차 렌더링 한다. 위 그림 3 메인 페이지의 소스를 보면 ``로 연결된 부분들이 링크 기능을 하는 부분이 된다.



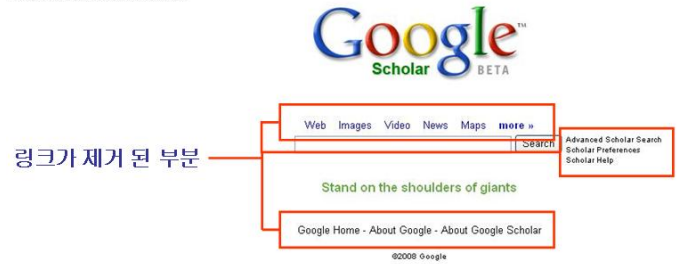
(그림 4) 1 차 렌더링 한 웹 페이지 소스와 하이퍼링크 부분

위 그림의 밑줄의 하이퍼링크의 내용들을 변경하게 된다. 그림 3 의 링크가 활성화 된 부분들이 모두 비활성화 된 것을 알 수 있다.

/test7.php

Test-MYH

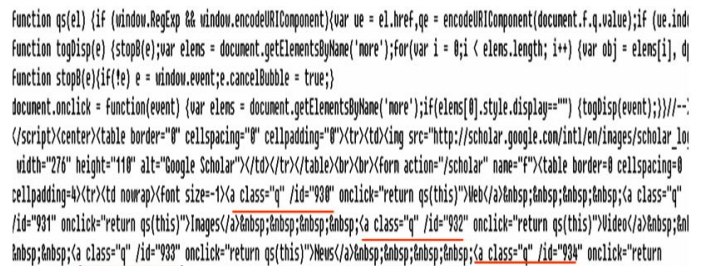
address inserted into database.



링크가 제거 된 부분

(그림 5) 1 차 렌더링 한 메인페이지

위의 웹 페이지는 링크 맵으로 한번 렌더링 한 웹 페이지이다. 위 그림에 표시된 사용자가 사용하지 않는 링크에 대해서 변환을 한다. 링크의 기능이 아직 활성화 된 상태로서 test7.php 파일로 변환 된 상태이다. 소스 부분의 링크 정보가 ID 태그로 변환함과 동시에 데이터베이스의 ID 정보가 들어가면서 그림 5 의 하이퍼링크는 비활성화 하게 된다. 그 동작을 구현한 후 결과는 다음과 같다.



(그림 6) 1 차 렌더링 메인 페이지의 소스 일부분

위 그림은 1 차 렌더링 한 웹 페이지의 소스 일부분을 보여주고 있다. 그림 4 의 하이퍼링크 정보부분이 변경된 것을 알 수 있다. 하이퍼링크의 태그 href 태그가 비활성화 되기 위해 id 태그로 변경되었으며 태그정보 역시 링크 주소가 아닌 링크주소가 저장된 데이터베이스 테이블의 아이디로 대체 되었다.

데이터 베이스에는 링크 id 가 자동으로 저장되어지며 링크 주소는 링크주소가 linkadd 로 들어가게 된다. 그리고 링크 주소가 하는 기능을 구별하기 위해서 링크 주소의 기능을 linkname 으로 하고 링크의 이름을 넣어 주게 된다.

linkid	linkadd	linkname
947	http://scholar.google.com/intl/en/scholar/about.ht...	About Google Scholar
946	http://www.google.com/intl/en/about.html?target=...	About Google
945	http://www.google.com/webhp?hl=en&lr="" target="" to...	Google Home
942	http://scholar.google.com/advanced_scholar_search?...	Advanced Scholar Search
943	http://scholar.google.com/scholar_preferences?hl=e...	Scholar Preferences
944	http://scholar.google.com/intl/en/scholar/help.htm...	Scholar Help
941	http://scholar.google.com/intl/en/options/? class=...	even more »
940	http://www.google.com/ptshp?tab=st" onclick="retur...	Patents
939	http://groups.google.com/grhp?tab=sg" onclick="re...	Groups
938	http://www.google.com/prdhp?tab=sf" onclick="retur...	Products
937	http://books.google.com/bkshp?tab=sp" onclick="ret...	Books
935	http://scholar.google.com/intl/en/options/? class=...	more »
936	http://blogsearch.google.com/?tab=sb" onclick="ret...	Blogs
934	http://maps.google.com/maps?tab=sl" onclick="retur...	Maps
933	http://news.google.com/nwshp?tab=sn" onclick="retu...	News
932	http://video.google.com/?tab=sv" onclick="return q...	Video
931	http://images.google.com/imghp?tab=si" onclick="re...	Images
930	http://www.google.com/webhp?tab=sw" onclick="retur...	Web

(그림 7) 링크 주소와 정보가 저장된 데이터 베이스

위 그림 7은 웹 페이지 소스에서 추출된 링크 정보들이 데이터 베이스에 저장된 것을 보여주고 있다. 이렇게 추출되어 저장된 링크 정보는 추후 사용자가 필요로 하는 링크 만을 제공하는 이용된다.

그림 4 와 그림 6 의 웹 페이지의 변경 전 웹 페이지 소스와 변경 후의 웹 페이지 소스는 그 사이즈에도 변화가 생긴다. 위 데이터 베이스에 저장된 링크 주소들을 보면 알 수 있듯이 긴 링크 주소의 문자열들이 짧은 링크 id 로 대체 되면서 웹 페이지의 소스 역시 4KB 에서 3.55KB 로 줄어 들었다. 위의 샘플 웹 페이지의 소스가 그리 큰 사이즈의 파일이 아니기 때문에 큰 차이를 느낄 수 없을 수도 있지만 서비스가 여러 개 연결된 페이지의 경우 그 차이를 더욱 쉽게 알 수 있다.

그림 5 의 경우를 보면 사용자가 바로 검색을 할 경우 그 외의 링크 기능을 비활성하고 서비스의 여부만을 보여주게 된다. 사용자가 검색만 할 경우 굳이 링크 기능을 모두 활성화 할 필요는 없기 때문에 전혀 검색하는데 있어서 문제가 발생하지 않는다.

검색 사용자가 필요로 하는 검색 기능 부분만을 활성화 함으로써 사용자의 메인 웹 페이지의 로딩을 빠르게 하고 동시에 다수의 사용자가 더 많이 접근 할 수 있게 된다.

5. 결론 및 향후 연구

현재 수 많은 상업적 사용자들을 위한 포탈 사이트 및 기업 사이트들이 서비스를 하고 있다. 다양하고 폭넓은 사용자들을 수용하기 위해 웹 서비스를 하는 사이트는 점점 그 구조가 복잡해지고 서비스의 양도 기하급수적으로 늘어나고 있다. 하지만 이런 서비스의 양적 증가는 사용자들에게 불필요한 서비스들의 제공으로 정작 원하는 서비스를 제공하는 저해요인이 될 수 있다. 이런 문제점들을 보완하고자 링크 맵 웹 시스템을 제안하였

다. 이 시스템의 사용을 통해서 사용자에게 대해 적절한 서비스를 제공할 뿐만 아니라 대부분의 웹 페이지의 서비스들은 수많은 링크들로 제공되기 때문에 필요한 링크만 기능하도록 활성화 시킴으로써 초기 웹 페이지의 사이즈의 크기를 줄일 수 있다. 이것은 초기 웹 페이지의 로딩을 좀 더 빠르게 할 수 있으며 많은 사용자들이 동시에 접근 할 경우 좀 더 많은 사용자들에게 서비스 할 수 있다.

위의 구현된 링크 맵 웹 페이지 시스템을 이용하여 여러 종류의 웹 서비스 페이지들에 적용하여 기존의 웹서비스 페이지와의 비교 연구가 더 필요하다. 또한 더 나아가 사용자들에 대한 서비스는 어떤 기준으로 이루어져야 하는가에 대해서도 연구가 필요하다. 사용자들에 대한 클러스터링을 한 후 각 클러스터링 된 그룹에 맞는 링크 맵을 만들어줘야 할 것이다. 현재 수많은 사용자의 링크 접근 패턴을 찾아내고 그 정보를 저장하는데 적지 않은 어려움이 있을 수 있지만 효율적인 링크 맵을 제공하여 사용자들에게 좀 더 편리한 서비스를 제공할 수 있을 것이다.

6. 참고문헌

[1] Yu Bo, Qi Luo. "Personalized Web Information Recommendation Algorithm Based on Support Vector machine", International Conference on Intelligent Pervasive Computing, pp487-pp490, 2007

[2] Lan Nie, Brian D. Davison, Xiaoguang Qi. "Topical Link Analysis for Web Search", SIGIR04, Agust 6-11, pp91-pp98, 2006,

[3] J.M. Kleinberg. Authoritative sources in a hyperlinked environment. Journal of the ACM, 46(5):604-632,1999

[4] 이수철, 이시은, 황인준, "웹 접근성 향상을 위한 웹 서핑 도우미", 정보과학회논문지 소프트웨어 및 응용 제31권 제9호, 2004.9

[5] 정은옥, 여정모, "EMFG를 이용한 쇼퍼몰 사이트 링크 구조에 관한 연구", 한국정보처리학회 추계학술대회 논문집 제10권 제2호, 2003.11

[6] 이원중, 정은재, 주수중, 이승용. "웹 통합문서의 효율적 생성과 검색을 위한 자동링크지원 시스템의 설계 및 구축", 정보처리학회논문지 A 제 10 권-A 권 제 2 호, 2003

[7] 염미령, 홍일구. "작업의 잔여량을 이용한 웹서버 로드 밸런싱", 정보처리학회 추계학술발표논문집, 제 8 권 제 2 호, 2001