

효율적인 온라인 신문기사 검색 브라우저

조이기^o 황인문 이성재 김재갑 김원중
순천대학교 컴퓨터과학과
{lgcho, him, sjlee, kjg, kwj}@cs.sunchon.ac.kr

A Browser for the Efficient On-Line Newspaper Search

Lee-Gi Cho^o In-Moon Hwang Sung-Jae Le Jae-Gab Kim Won-Jung Kim
Dept. of Computer Science, Sunchon National University

요 약

신문은 가장 오래된 정보분배 수단인 하나이고, 대부분의 사람들은 방송보다는 좀 더 자세한 일상의 정보를 획득하기 위하여 날마다 신문을 이용한다. 인터넷의 등장과 함께 모든 주요 신문사들은 온라인 신문기사를 독자들에게 제공하고 있으며, 신문사들의 웹 사이트들은 페이지 뷰(Page View)의 평가에서 대부분이 상위에 랭크되어 있다. 그러나 현재의 온라인 신문기사의 브라우저에는 사용자의 이용 편의성 및 시간에서 많은 문제점을 내포하고 있다. 본 논문에서는 사용자가 온라인 신문기사를 검색할 때, 신문기사 전체의 윤곽을 파악할 수 있도록 하고, 브라우저에 소요되는 시간과 마우스 클릭 횟수를 대폭 줄일 수 있는 브라우저 방법에 대해 연구하였다.

1. 서론

신문(Newspaper)은 인류역사에서 가장 오래된 정보분배 수단의 하나이고, 현재에도 다른 미디어와 비교하여 방송 다음으로 가장 많은 사람들이 이용하고 있다. 시청각을 이용하는 방송이 약간의 수동적인 입장에서의 정보획득 수단이라면, 신문은 독자 스스로가 신문을 집어들어 펼쳐야만 볼 수 있는 매우 능동적인 정보획득 요구를 필요로 한다. 즉, 사람들은 일상생활 또는 사회생활에 필요한 그러면서도 방송에서의 보도보다는 좀 더 자세한 정보를 획득하기 위하여 날마다 신문을 이용한다. 또한 신문은 동시에 많은 기사를 함께 표현하여 짧은 시간에 많은 기사를 읽어야 하는 바쁜 독자들을 위한 훌륭한 표현 기법을 가지고 있다. 각 신문사마다 수백 명의 기자들이 현장의 생생한 목소리를 전달하며, 현재의 인터넷에서 해결하여야 할 핵심 기술로 부상하고 있는 정

보분류(Information Clustering)를 가장 정확히 표현하고 있는 매체이다. 즉, 우리는 신문의 머리기사, 정치, 경제, 사회, 문화의 기사가 어디에 어떻게 표현되는지를 이미 알고 있다. 인터넷의 등장과 함께 모든 주요 신문사들은 온라인 신문기사를 독자들에게 제공하고 있으며, 많은 사람들이 웹에서 여러 종류의 신문 기사를 실시간으로 검색하고 있다. 또한 각 신문사들이 상대적으로 높은 페이지 뷰(Page View)를 바탕으로 신문기사뿐만 아니라 금융, 쇼핑, 교육, 정보검색, 메일링, 동호회 등 기존 IT 업계의 전문영역까지도 포함하는 포털 사이트로 발전하고 있어, 각 신문사의 웹 사이트는 더욱 더 많은 이용이 예견되고 있다. 그러나 현재 온라인 신문기사의 브라우저는 독자의 이용 편의성 및 소요되는 브라우저 시간에서 몇 가지 문제점을 내포하고 있다.

본 논문에서는 사용자가 온라인 신문기사를 검색할 때, 신문기사 전체의 윤곽을 파악할 수 있도록 하고,

브라우저에 소요되는 시간과 마우스 클릭 횟수를 대폭 줄일 수 있는 브라우저 방법에 대해 연구하였다.

2. 현재 온라인 신문기사 검색의 문제점

인터넷 포털사이트(Internet Portal Site)는 네티즌들이 인터넷 접속을 위해 처음으로 방문하는 출입 사이트로 인터넷상에서 가능한 모든 서비스와 콘텐츠를 종합적으로 제공하며, 정보를 찾기 위한 길잡이 역할을 수행한다. 본 논문에서는 이와 유사하게 웹 포털페이지(Web portal Page)를 정의한다. 즉, 포털페이지는 어떤 웹 사이트의 정보를 검색하기 위한 출입구로서 사용자는 이 페이지를 반복하여 가장 빈번하게 접근하게 된다. 신문사의 웹 사이트에서는 처음 접속하면 나타나는 기사들의 리스트를 포함하고 있는 페이지가 이에 해당한다. 현재 온라인 신문기사 브라우저의 문제점은 다음과 같다.

2.1 온라인 신문기사 브라우저

사용자들은 온라인 신문기사를 브라우저하기 위하여 해당 신문사의 포털페이지에 접속한다. 먼저 주요 기사의 리스트를 살펴 본 후에 본인이 관심을 가지고 있는 기사의 타이틀을 클릭하여 해당 기사를 읽게 된다. 기사 내용을 읽은 후, 다음의 원하는 기사를 읽기 위하여 다시 포털페이지로 돌아 와야만 한다. 이 때, 본인이 기사를 읽은 계층의 수만큼 'Back' 키를 사용한다. 따라서 신문기사 내용을 브라우저 하는 동안에는 전체기사의 윤곽을 파악할 수 없고, 불필요한 많은 마우스 클릭이 필요하게 된다. 그러므로 이러한 문제점을 해결하면 훨씬 효율적으로 온라인 신문기사를 검색할 수 있을 것이다. 본 연구에서는 다중윈도우(Multi-windows)를 사용하여 이 문제를 해결하였다 [3].

2.2 프락시 웹 서버

프락시 웹 서버(Proxy Web Server)는 네트워크의 트래픽량과 정보를 제공하는 서버에서 응답하여야 하는 클라이언트의 요청(Request) 개수를 줄여, 서비스의 시간성능을 개선하기 위하여 일반적으로 사용되고 있다. 프락시 웹서버는 다음과 같은 순서로 수행된다[2].

- ① 웹 페이지에 대한 사용자의 요청을 인터셉트(Intercept) 한다.
- ② 요청된 페이지의 복사본이 로컬 캐시에 존재하는지 체크한다.
- ③ 만일 그것이 로컬 캐시에 존재하고 원래의 서버에

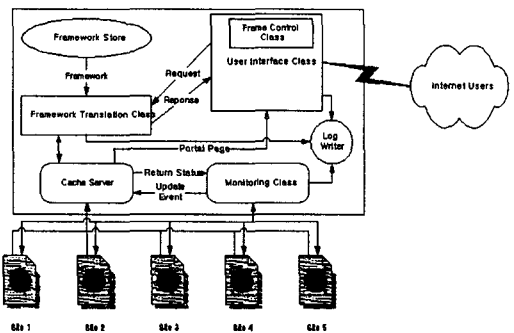
서 변경되지 않았다면, 그 페이지를 사용자에게 리턴 한다.

④ 만일 로컬 캐시에 존재하지 않거나, 존재하지만 원래 서버에 있는 내용과 다르면(변경되었으면), 원래 서버에서 그 내용을 가져와 로컬 캐시에 저장한 후, 그 내용을 사용자에게 리턴 한다.

이러한 프락시 웹 서버의 수행 시나리오는 온라인 신문기사 검색의 경우에 획기적으로 개선될 수 있다. 왜냐하면 '온라인 신문 검색의 경우 검색하는 기사의 개수만큼 포털페이지를 접근하기 때문이다. 즉, 온라인 신문기사를 검색할 때는 포털페이지에 대해서는 로컬 캐싱을 이용하지 않고 주기적장치에 상주시켜 관리하면, 검색속도를 크게 높일 수 있다.

3. 효율적인 신문기사 검색 브라우저

본 논문에서 구현한 효율적인 온라인 신문검색 브라우저를 위한 시스템은 윈도우2000에서 자바 어플리케이션과 서블릿(Servlet)을 사용하여 인터넷익스플로러 5.x 상에서 개발하였으며, 개발도구로는 JDK 1.2, OS/2K 2.0를 이용하였다. 시스템은 [그림1]과 같은 구조를 가지고 있으며, 주요 모듈의 기능은 다음과 같다.



[그림 1] 시스템 구조

(1) User Interface Class

User Interface Class 모듈은 사용자와 시스템간의 인터페이스를 지원하는 모듈로서 [그림2]와 같이 3개의 프레임 구조로 이루어져 있다. 왼쪽 위의 프레임은 사용자가 읽기를 원하는 신문을 선택하는 창이고, 아래쪽의 프레임은 해당 온라인 신문의 포털페이지를 디스플레이 하는 창이다. 오른쪽 프레임은 사용자가 신문기사를 선택하였을 때, 그 기사내용을 디스플레이

하는 창이다. 이 모듈은 사용자가 읽기를 원하는 온라인 신문을 선택하면, 그 내용을 Framework Translation Class로 전송하고 사용자의 세션정보를 유지한다. 그리고 Framework Translation Class의 수행 결과 해당 포털페이지의 프레임 구조에 변화가 없으면, 캐시서버로부터 해당 신문의 포털페이지를 가져와 사용자에게 디스플레이 한다. 만일 프레임의 구조가 바뀌었다면(매우 드문 경우이지만) 본 시스템에서는 현재 우리가 사용하는 원래의 브라우징 방법으로 돌아가도록 하였다. 이 경우 본 시스템을 다시 사용하도록 하려면 Framework Store 부분을 변경하여 재구축 하여야 한다.

(2) Monitoring Class

Monitoring Class는 시스템에서 유지관리하고 있는 온라인 신문 사이트들을 주기적으로 순회하여 사이트의 포털 페이지에 변동이 발생하는 경우 캐시서버에 저장된 해당 사이트의 포털 페이지를 갱신한다. 본 연구에서는 한겨레 신문, 조선, 한국, 중앙, 한국일보 5개 신문의 온라인 사이트를 대상으로 시스템을 구축하였다. 실험결과 네트워크가 정상적인 경우 5개 사이트의 모든 포털 페이지를 다운로드하여 캐시 서버에 저장하는 시간은 3분을 초과하지 않았다. 그러나 최적의 순회시간을 산출하기 위해서는 네트워크가 비

(3) Framework Store

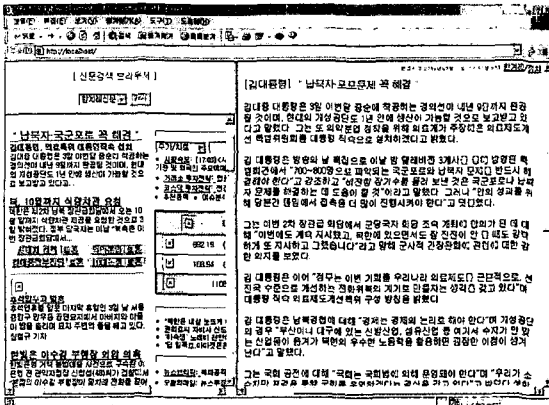
Framework Store는 특정 도메인 내의 어플리케이션을 위해 확장 가능한 템플릿을 제공하는 아키텍처 패턴으로 각 사이트의 포털페이지의 구성에 대한 규칙 정보를 가지고 있다. 온라인으로 기사를 제공하는 신문 사이트는 Framework를 기반으로 운영된다. 따라서 구현한 시스템의 Framework Store는 각 온라인 신문 사이트들의 포털페이지에 대한 Framework 정보를 보관하고 있다. 본 연구에서 구현한 온라인 신문사들의 신문기사 리스트에 대한 참조링크 형식은 [표1]과 같았다. Framework Translation Class는 이 정보를 이용하여 캐시서버에 저장된 포털페이지의 HTML 문서를 구현 시스템의 사용자 인터페이스에서 디스플레이 가능한 형태로 변환시킨다.

구분	신문기사에 대한 링크 구조
한국일보	href=/NewsPortal/
한겨레	href="/section-
동아일보	HREF="/fbin/
조선일보	href="/w21data/, href=http://www.chosun.com/w21data/
중앙일보	href="/news/

[표 1] 신문기사에 관한 링크 정보

(4) Framework Translation Class

Framework Translation Class는 User Interface Class로부터 사용자의 신문 선택에 관한 정보가 넘어오면, 캐시서버에 저장된 포털페이지의 HTML 문서와 Framework Store에 저장된 해당 사이트의 프레임 정보를 이용하여 사용자 인터페이스 화면에 디스플레이 가능한 형태로 변환시킨다. 아래의 [표2]와 [표3]은 한겨레 신문에서 원래의 포털페이지와 변환된 페이지 내용의 일부를 보여 주고 있다. 구현된 시스템은 효율적인 신문기사 브라우징을 위한 시스템이기 때문에 포털페이지에서 신문기사가 아닌 광고나 다른 URL을 클릭하면 구현된 시스템의 사용자 인터페이스 화면을 무시하고 target="_parent"를 이용하여 현재 우리가 사용하고 있는 브라우징 형태와 동일하게 디스플레이 하게 된다.



[그림 2] 구현 시스템

정상적인 경우를 고려하여야 한다. 구축된 시스템에서는 5번까지 사이트에 접근을 시도한 후에도 실패하면, Log Writer에 해당 정보를 저장한 후, 캐시서버의 내용을 갱신하지 않고 다음 사이트로 넘어 가도록 하였다.

```

* 신문기사인 경우
<a href="/section-003000000/2000/003000000200009031857005.html">
* 신문기사가 아닌 경우
<a href=http://www.hani.co.kr/section-special/cbh.html>

```

[표 2] 원래의 HTML 페이지

```

* 신문기사인 경우
<a href="/section-003000000/2000/003000000200009031857005.html"
target="contents">
* 신문기사가 아닌 경우
<a href=
http://www.hani.co.kr/section-special/cbh.html target="_parent">

```

[표 3] 변경된 HTML 페이지

10th annual ACM Symposium on User Interface Software and Technology, 1997.

- [4] <http://www.donga.com/>
- [5] <http://www.hani.co.kr/>
- [6] <http://www.joins.com/>
- [7] <http://www.chosun.com/>
- [8] <http://www.hankooki.com/>

4. 결론

현재 온라인 신문기사의 브라우징은 사용자가 전체 기사의 윤곽을 항상 지켜보면서 작업할 수 없으며, 포털페이지를 중심으로 전체 마우스 클릭의 1/2은 전혀 불필요하다. 본 논문에서는 이러한 문제점을 해결하기 위하여 효율적인 신문기사 검색 브라우저를 위한 보조 시스템을 개발하였다. 구현된 시스템을 이용하면 기사를 브라우징하는 동안에 계속하여 전체 기사의 윤곽을 파악할 수 있으며, 하나의 기사를 읽은 후에 다음 신문기사를 읽기 위하여 'Back' 키를 전혀 클릭할 필요가 없기 때문에 기존의 브라우징 기법을 이용할 때 보다 최대 1/2까지 마우스 클릭 횟수를 줄일 수 있다.

향후 연구과제로는 신문기사뿐만 아니라 유용한 온라인 신문 정보를 이용하여 개인 맞춤형 브라우징 시스템을 개발하는 것이다.

5. 참고 문헌

- [1] C.M. Bowman et al, "Customizable Discovery and Access System," Tech. Rep. CU-CS-732-94, Dept. Comp. Sci., Univ. Colorado-Boulder, 1994.
- [2] D. O'Callaghan, "A Central Caching Proxy Server for WWW Users at the University of Melbourne," Proc. AusWeb '95, Ballina, New South Wales, Australia, 1995.
- [3] E. Kandogan and B. Shneiderman, "Elastic Windows: A Hierarchical Multi-Window World-Wide Web Browser," Proceedings of the