

웹서버 로그를 이용한 웹 콘텐츠 사용료 정산 방안 연구

조충래, 김용호, 남기서
한국기계연구원 기계재료기술정보센터

A Study of Estimating the Usage of Web Contents Using the Logs of Web Server

Chunglae Cho, Yongho Kim, Giseo Nam
Information Center of Machinery & Materials, Korea Institute of Machinery & Materials

요 약

콘텐츠 산업의 증가와 함께 웹을 기반으로 하는 콘텐츠 제공 산업 또한 크게 증가하고 있다. 이에 따라 웹 콘텐츠에 대한 효율적이고 합리적인 사용료 정산 방법이 요구된다. 웹을 기반으로 하는 콘텐츠 서비스의 경우 웹의 특성상 서버와 클라이언트의 연결이 지속되지 않으므로 사용 시간에 따라 사용료를 부과하기가 어렵다. 웹서버의 로그를 이용한 콘텐츠 사용료 부과 방식은 사용자가 사용한 콘텐츠의 질과 양에 따라 사용료를 부과할 수 있는 효과적인 방법이다. 콘텐츠 사용자 정보와 콘텐츠 사용료 부과기준 정보를 이용하여 웹서버의 로그 파일을 분석하여 사용자별 콘텐츠 사용료를 구한다. 웹서버 로그를 이용한 콘텐츠 사용료 정산 방법은 모든 형태의 웹 콘텐츠 제공 시스템에 쉽게 적용할 수 있다.

1. 서 론

최근 콘텐츠 산업의 증가와 더불어 멀티미디어 콘텐츠 산업이 급속도로 발전하고 있다. 멀티미디어 콘텐츠란 문자, 소리, 화상, 영상 등의 콘텐츠 정보가 디지털로 제공되고, 정보 수용자와 제공자 사이의 양방향성을 보장해 주는 정보를 의미한다[1,3]

최근 몇 년 사이 웹이 정보 산업의 인프라로 급 부상하면서, 웹을 기반으로 하는 콘텐츠 제공 산업 또한 크게 증가하고 있다. 웹을 통해 제공되는 콘텐츠로는 전자 출판물, 영상, 오락 및 게임 프로그램, 정보 서비스 및 데이터베이스 등을 포함한다[1].

웹 콘텐츠의 사용료 정산 문제는 웹 콘텐츠 서비스의 발전과 더불어 해결해야 할 가장 기본적인 문제중 하나이다. 인터넷이 아닌 공중망을 이용한 텍스트 위주의 콘텐츠 서비스의 경우에는 사용 시간을 근거로 사용료를 부과하는 경우가 대부분이었다. 하지만, 웹을 통해 콘텐츠 제공 서비스를 할 경우에는 웹의 특성상 사용 시간을 구하기가 힘들며, 사용자의 위치와 사용 시간에 따라서 인터넷 트래픽의 차이가 심하기 때문에 시간 요소로 요금을 부과한다는 것은 형평성에 어긋난다.

본 논문에서는 효율적인 웹 콘텐츠 사용료 부과를 위해서 웹서버의 로그 파일을 이용하는 방법을 제안한다.

2. 연구배경

웹 콘텐츠 산업의 발달과 함께 유료 콘텐츠 사이트 또한 증가하고 있지만, 사용료의 정산 방법이 만족스럽지 못한 경우가 많다. 대부분 콘텐츠의 사용량에 상관없는 정액제를 채택하고 있기 때문이다. 하지만 정액제는 사용자의 이용량에 상관없이 일정한 요금을 부과하는 방식이므로 서비스를 적게 사용하는 사용자들이 상대적으로 불만을 가질 수 있다. 또한 제공 콘텐츠의 중요도에 상관없이 일정한 요금을 부과하게 되므로 콘텐츠 제공자 입장에서 만족스럽지 못한 요금 정산 방법이다.

콘텐츠 사용자와 제공자 모두를 만족시킬 수 있는 요금 부과 방식은 바로 콘텐츠를 사용한 만큼의 사용료를 부과하는 콘텐츠 단위 종량제 방식이다. 콘텐츠 단위 종량제 방식은 사용자가 접근한 콘텐츠 별로 요금을 부과하는 방식으로, 콘텐츠의 중요도에 따라 사용료를 차별하여 부과할 수 있으며, 사용량에 따라 요금을 달리 책정할 수 있는 방법을 말한다.

본 논문에서는 이러한 콘텐츠 단위 종량제를 쉽게 적용할 수 있는 방법으로 웹서버의 로그를 이용하여 사용료를 정산하는 방법을 제안한다.

3. 웹 콘텐츠 사용료 부과 시 고려 사항

웹서버 로그를 이용한 사용자별 정산 과정은 크게 두 부분으로 나뉜다.

한 부분은 웹서버가 사용자의 요구를 콘텐츠 서버에 전달하고 콘텐츠 서버가 제공하는 콘텐츠를 사용자에게 전달해 주는 부분이다. 이 과정에서 웹서버는 사용자가 콘텐츠 접근이 허용된 사용자인지를 검증하며 사용자가 콘텐츠를 사용한 현황을 웹서버 로그에 기록한다.

또 한 부분은 웹 서버가 만든 웹 로그 파일을 사용하여 사용자별 콘텐츠 사용료를 정산하는 콘텐츠 사용료 정산 서버 부분이다. 콘텐츠 사용료 정산 서버는 웹서버 로그에 남겨진 사용자의 콘텐츠 사용 내역을 바탕으로 콘텐츠 사용자 정보와 콘텐츠 사용료 부과 기준 정보를 이용하여 사용자별 콘텐츠 사용료 데이터 및 각종 콘텐츠 사용 현황 분석 데이터를 생성한다.

웹서버 로그 파일이 생성되는 과정은 그림 3과 같다. 사용자는 콘텐츠를 액세스하기 위해서 자신의 아이디와 패스워드를 웹서버에 전달한다. 웹서버는 웹 사용자 정보를 참조하여 허용된 사용자인지를 검증한다. 접근이 허용된 사용자일 경우 사용자는 콘텐츠가 담긴 웹 페이지들을 서핑하며 유료 콘텐츠를 검색할 수 있다. 이러한 모든 행위들이 웹서버 로그 파일에 기록된다.

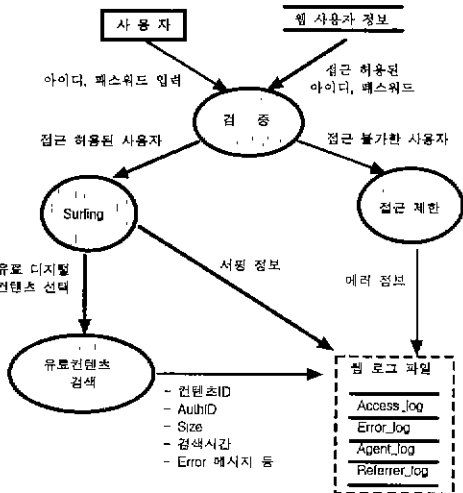


그림 3. 웹서버 로그 생성 흐름도

웹서버 로그를 이용하여 콘텐츠 사용료를 정산하는 과정은 그림 4와 같다. 먼저 로그 분석 모듈이 웹서버 로그를 분석하여 콘텐츠 액세스와 관련된 로그들만을 따로 필터링한 로그 데이터를 생성한다. 그러면 사용료 정산 모듈은 콘텐츠 사용자 상세 정보와 콘텐츠 사용료 부과기준 정보를 가지고 사용자별 콘텐츠 사용료를 구할 수 있다. 사용자별 콘텐츠 사용료 데이터는 리포트 생성 모듈에 의해서 사용자별 사용료 정산 리포트와 콘텐츠별 사용 현황 분석 리포트를 생성하는데 이용된다.

6. 결론 및 향후 연구 과제

웹서버의 로그를 이용한 콘텐츠 사용료 부과 방식은 사용자가 사용한 콘텐츠의 질과 양에 따라 사용료를 부과하는 합리적이고 쉬운

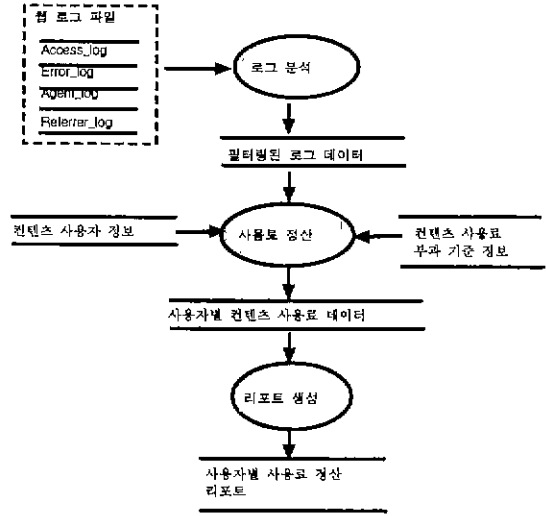


그림 4. 콘텐츠 사용료 정산 흐름도

사용료 정산 방법이다. 이 방법은 웹을 통해서 서비스하는 모든 형태의 콘텐츠 제공 시스템에 적용할 수 있다. 또한, 이 방식의 콘텐츠 사용료 정산 시스템은 콘텐츠 제공 서버와 완전히 분리될 수 있기 때문에 기존의 웹 콘텐츠 제공 시스템에 쉽게 포팅이 가능하다.

콘텐츠 사용료 부과 기준 데이터를 어떻게 관리하느냐에 따라 콘텐츠 제공자가 원하는 다양한 형태의 사용료 정산을 할 수 있다. 앞으로 이에 대한 심도 있는 연구가 요구된다.

【참고 문헌】

- [1] 김도환, 박성진, 김유정, "콘텐츠", 정보통신산업동향, 정보통신연구원, 1997년 12월 15일
- [2] 배성민, "WWW Build, Refine and Management", WWW Workshop, 제4회, 제1권, p.51-75, 1996년 11월
- [3] "인터넷(NIC) 동향 및 기술연구", 한국전산원 연구보고서, 1997년 12월
- [4] Ari Luotonen, Tim Berners-Lee, "CERN httpd User's Guide/Reference Manual," CERN, URL: <http://www.w3.org/Daemon/User/Guide.html>
- [5] "Internet Product Watch for a List of Analysis Tools", URL: <http://ipw.internet.com/analysis/index.html>
- [6] Ming-Syan Chen, Jong Soo Park, Yu, P.S., "Data Mining for Path Traversal Patterns in a Web Environment", Proceedings of the 16th International Conference on Distributed Computing Systems, p.385-392, May 1996
- [7] "Web Servers Survey", URL: <http://www.webcompare.com/server-survey.html>
- [8] Zaane, O.R., Man Xin, Jiawei Han, "Discovering Web Access Patterns and Trends by Applying OLAP and data mining technology on Web Logs", Proceedings. IEEE International Forum on Research and Technology, Advances in Digital Libraries, p.19-29, April 1998

앞에서 언급한 바와 같이 웹 콘텐츠 서비스의 경우 사용자 부과 기준을 사용 시간으로 정할 경우 다음과 같은 문제점이 있다.

첫째, 웹의 특성상 서버와 클라이언트의 연결이 지속되지 않기에 때문에 사용자가 특정 웹 페이지를 사용한 시간을 알 수 없다 이러한 문제를 해결하기 위한 연구가 진행되고 있지만 아직까지는 해결되지 않은 상태이다[2].

둘째, 인터넷의 특성상 모든 사용자가 같은 조건에 의해 콘텐츠를 접근하지 못하므로, 콘텐츠 사용시간은 합리적인 사용료의 기준이 될 수 없다. 사용자의 위치에 따라 네트워크 속도가 다르며, 사용 시간에 따라 네트워크 트래픽이 다르기 때문에 사용 시간에 의해 콘텐츠 사용료를 부과하는 것은 공평한 방법이 아니다.

웹을 기반으로 하는 콘텐츠 서비스의 경우 콘텐츠 사용료 부과 시 다음과 같은 고려 사항들이 있을 수 있다.

첫째, 사용자가 접근한 콘텐츠의 질에 따라 사용료를 차등 부과 할 수 있어야 한다. 즉, 모든 콘텐츠가 똑같은 가치를 지니지는 않으므로 콘텐츠의 질에 따라 사용료를 달리 부과할 수 있어야 한다.

둘째, 사용자가 접근한 콘텐츠의 량에 따라 사용료를 차등하여 부과할 수 있어야 한다.

셋째, 일정 시간 내의 같은 콘텐츠 접근은 추가 사용료를 부과하지 않아야 한다. 인터넷의 특성상 모든 사용자가 같은 조건(네트워크 속도, 네트워크 트래픽)에서 서비스를 제공받지 못한다. 어떤 사용자의 경우는 연결 상태가 좋지 않아서 다시 접근해야 하는 경우도 있다. 그런 경우 접근할 때마다 사용료를 부과한다면 사용자가 불만을 가질 것이다. 또한 하이퍼텍스트 형태인 웹의 특성상 이미 접근한 콘텐츠를 다시 접근하는 경우가 많기 때문에 일정 시간 내의 같은 콘텐츠에 대한 접근은 한 번의 접근으로 보아야 한다.

넷째, 사용료 부과에 관련된 모든 변수는 서비스 제공 특성에 따라 조정 가능해야 한다.

본 논문에서 제안하는 웹서버의 로그를 이용한 사용료 정산 방법은 위의 고려 사항을 모두 만족하는 방법이다.

4. 웹서버 로그를 이용한 콘텐츠 사용료 정산의 타당성

웹서버의 로그는 웹의 사용자가 서버 쪽에 접근한 내용을 기록한 파일이다. 웹서버의 로그에는 사용자 아이디, 사용자가 접근한 시간, 사용자의 IP 어드레스, 사용자가 접근한 파일(즉, 콘텐츠), 사용자가 접근한 량 등이 기록되어진다(그림 1)[4,7]. 웹서버의 로그는 웹의 트래픽 분석이나 사용자의 패턴 분석을 위한 데이터로 활용되고 있다 [5,6,8].

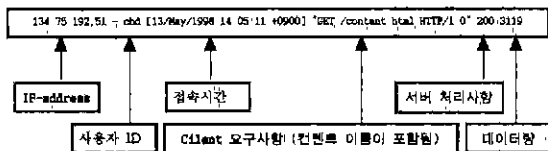


그림 1. 웹서버 로그 파일의 예

웹서버의 로그를 콘텐츠 사용료 부과에 이용할 수 있는 근거는 다

음과 같다.

웹을 통한 서비스는 반드시 웹서버가 필요하며, 웹 사용자가 웹 콘텐츠로 접근하고자 할 때는 반드시 웹서버를 통해서만 가능하다. 웹 서버를 통해서 서버 내의 자원에 접근하는 모든 상황은 웹서버 로그에 기록되므로, 어떤 사용자가 어떤 콘텐츠를 접근했는지에 대한 내용이 로그에 남게 된다. 따라서, 이 로그 파일을 분석하여 콘텐츠 사용료 정산의 데이터로 활용할 수 있는 것이다.

5. 웹서버 로그를 이용한 콘텐츠 사용료 정산 시스템

웹서버 로그를 이용한 사용료 정산 시스템의 구조는 그림 2와 같다. 각 서버 모듈의 기능 요약은 아래 표 1과 같다.

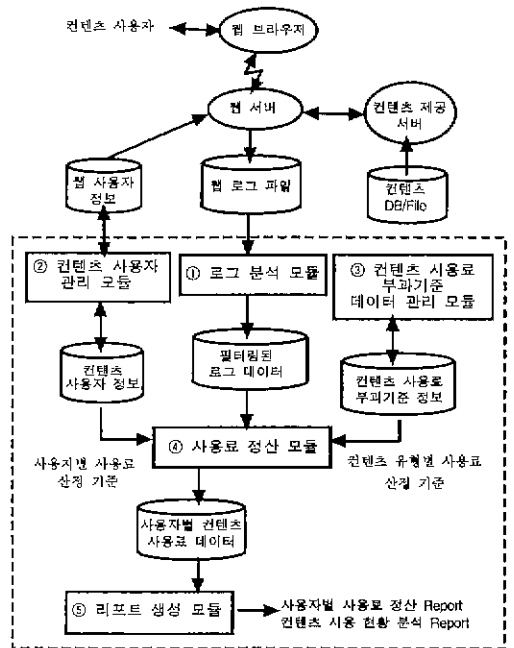


그림 2. 웹 콘텐츠 사용료 정산 시스템 전체 구성도

표 1. 서버 모듈 기능 요약

모듈명	기능요약
로그분석모듈	웹서버가 생성한 웹 로그를 분석하여 콘텐츠 액세스와 관련된 로그들만을 따로 필터링한 로그 데이터를 생성
콘텐츠 사용자 관리 모듈	콘텐츠를 사용할 수 있도록 인증된 사용자에 대한 정보를 관리
콘텐츠 사용료 부과 기준 데이터 관리 모듈	콘텐츠별로 사용료를 부과하기 위한 정보들을 관리
사용료 정산 모듈	필터링된 로그 데이터, 콘텐츠 사용자 정보, 콘텐츠 사용료 부과 기준 정보를 참조하여 사용자별 콘텐츠 사용료 데이터를 생성
리포트 생성 모듈	사용자별 사용료 정산 리포트와 콘텐츠별 사용 현황 분석 리포트 생성