

# 디지털 콘텐츠 거래 모니터링 시스템에서 로그 저장 장치의 안정성 증진 방안

민지홍\*, 송광호\*, 김유성\*  
인하대학교 정보통신공학과

e-mail:newindia89@gmail.com, crossofjc@gmail.com, yskim@inha.ac.kr

## A Method for Improving Stability of Log Storage in Digital Content Trade Monitoring System

Jihong Min\*, Kwangho Song\*, Yoo-Sung Kim\*

Dept. of Information and Communication Engineering, Inha University

### 요 약

본 논문에서는 디지털 콘텐츠 거래 모니터링 시스템의 로그 저장 장치의 안정성을 개선하는 방법을 제안한다. 디지털 콘텐츠 거래 모니터링 시스템은 디지털 콘텐츠의 거래 내용을 정확하게 파악하기 위해서 콘텐츠 거래 로그를 영속적으로 관리할 수 있어야 한다. 그러나 최근에 본 연구팀에 의해서 개발된 프로토타입 시스템에서는 대량의 트랜잭션이 실행될 경우에 저장 로그 데이터의 안정성이 훼손되는 경우가 발생할 수 있음이 보고되었다. 따라서 본 연구에서는 디지털 콘텐츠 거래 모니터링 시스템에 동시에 다수의 트랜잭션이 발생하여 검색엔진에 문제가 생기더라도 저장하고 있는 데이터의 안정성을 보장하면서 다수의 이용자들로부터 수집한 로그를 처리할 수 있도록 로그 저장시스템의 안정성 증진 방안을 제안한다.

### 1. 서론

한류, K-Pop 등으로 대표되는 한국의 디지털 콘텐츠 산업의 화려한 성장의 이면에는 저작권자(CP: Content Provider)와 온라인 서비스 유통자 (OSP : Online Service Provider)간의 불공정한 수익분배라는 문제가 존재한다[1]. 실제로 OSP들이 콘텐츠를 판매한 기록을 누락하기 위해 사용자들의 콘텐츠 거래 정보를 나타내는 로그기록을 조작하는 등의 일들이 벌어지고 있으며 [2], 이러한 문제들로 인한 OSP와 저작권자 간의 법정다툼도 벌어지고 있다 [3]. 따라서 이러한 갈등들이 디지털 콘텐츠 산업의 발전을 저해할 수 있다는 우려가 제기되고 있으며, 이에 대한 대책으로 저작권료의 공정한 분배 정산의 기초자료로 활용될 수 있는 거래 로그기록을 공신력 있는 기관이 수집 및 공개하여 투명한 거래 증빙 자료를 제공하자는 방법 등이 논의되고 있다 [1].

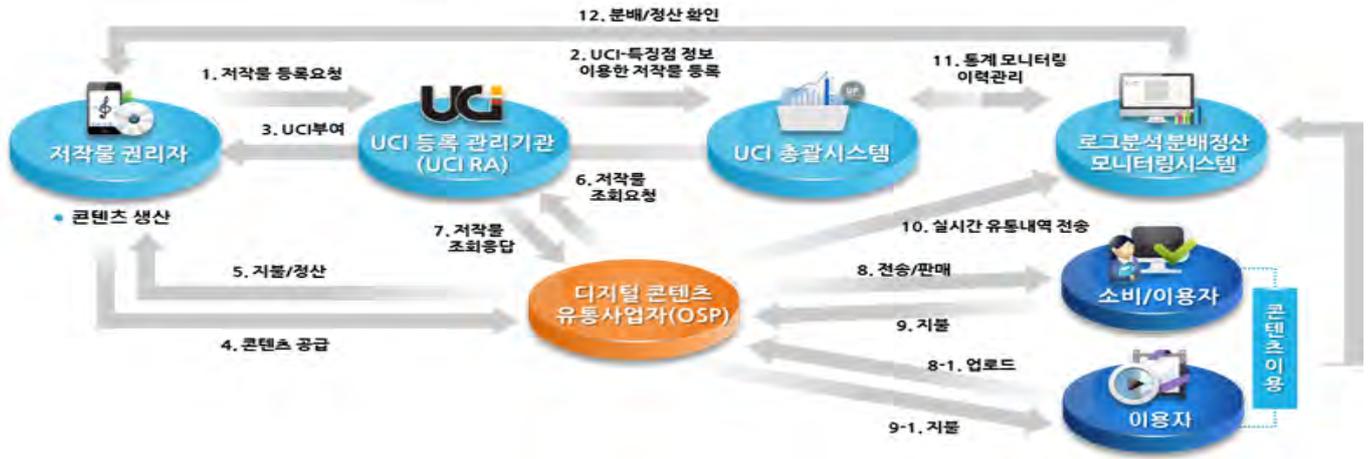
따라서 본 연구팀에서는 앞서 언급한 문제들을 해결하고 OSP가 수익의 투명한 분배정산을 하도록 하기 위해 국가표준식별체계인 UCI (Universal Content Identifier)를 기반으로 하는 디지털 콘텐츠 유통관리 서비스모델을 제안하였다 [4]. 또한 앞서 제안된 서비스 모델을 기반으로 디지털 콘텐츠 거래 모니터링 시스템의 핵심 모듈인 유통 로그 수집·분석 시스템을 설계하였고 [5], 이에 맞는 프로

토타입을 구현하였다 [6]. 이전 연구들에서 설계한 시스템이 저장하고 있는 로그데이터는 그 목적이 투명하고 공정한 분배정산의 기초자료로서의 활용인 만큼 삽입, 수정, 보관의 안정성이 매우 중요하다. 따라서 [4][5][6]에서 제안했던 시스템의 로그 데이터 저장소는 저장하고 있는 데이터에 대한 안정성을 필히 보장해야 한다. 그러나 추가적인 조사를 거친 결과 [6]의 프로토타입에 저장소로 사용한 Elasticsearch[7]는 과도한 질의 트랜잭션이 발생하면 Elasticsearch가 멈추면서 저장하고 있는 데이터의 손실이 발생하는 문제가 발생할 수 있어 저장한 로그 데이터의 안정성을 완벽하게 보장하지는 못한다는 문제점을 발견하게 되었다[8].

따라서 본 연구팀은 이를 해결하기 위해 본 논문에서는 [4][5][6]의 디지털 콘텐츠의 거래 모니터링을 위한 로그 수집·분석 시스템에서 로그 데이터를 안전하게 저장할 수 있는 로그 데이터의 안전한 저장소를 설계한다. 본 저장장치는 CP와 OSP간에 정확한 분배정산을 위해 필요한 거래 로그정보를 안전하게 저장 관리하기 위해 검색 엔진과 로그 저장소를 분리하여 설계하고 로그 저장소를 위해서는 대량의 트랜잭션을 효율적으로 처리하도록 설계된 관계형 데이터베이스 관리 시스템을 사용하여 저장한 로그 데이터의 안정성을 보장한다.

본 논문의 구성은 다음과 같다. 2절에서는 UCI 기반의 거래상황 모니터링을 위한 로그 수집·분석 시스템, 그리고 그 시스템이 갖춰야 할 특성들을 소개한다. 3절에서는 저

본 연구는 문화체육관광부 및 한국저작권위원회의 2015년도 저작권기술개발 사업의 연구결과로 수행되었음.



(그림 1) UCI 기반 디지털 콘텐츠 유통관리 서비스 모델의 흐름도

장하고 있는 로그데이터의 안정성을 보장하는 로그 데이터 저장장치를 제안한다. 마지막 4절에서는 본 연구의 결론 및 향후연구에 대해 기술한다.

## 2. 관련 연구

본 절에서는 UCI기반의 디지털 콘텐츠 유통관리 서비스모델 및 유통상황 모니터링을 위한 로그 수집·분석시스템, 그리고 그 프로토타입에 대해 기술한다.

### 2.1 UCI기반의 디지털 콘텐츠 유통관리 서비스 모델

UCI는 식별 가능한 자원의 효율적인 유통과 활용을 위하여 개별 자원에 유일한 코드를 부여하고 이를 관리해주는 식별체계이다 [1]. 본 연구팀은 동일 자원의 다른 카피들에 서로 다른 식별코드를 부여 및 활용할 수 있다는 점, 식별메타데이터의 identifier 필드를 이용해 다른 식별코드들과 매핑이 가능하다는 점, 응용 메타데이터의 항목 구성이 자유롭다는 점 등의 특징을 갖는 UCI를 거래되는 디지털 콘텐츠에 부착하여 디지털 콘텐츠의 유통 정보를 정확하게 수집하기 위해 UCI를 기반으로 하는 디지털 콘텐츠 유통관리 서비스 모델을 제안하였다 [4]. 제안한 서비스 모델은 아래 그림 1과 같은 흐름도를 가지며 디지털 콘텐츠 산업의 주체별 역할과 디지털 콘텐츠 거래의 투명성과 공정성을 확보할 수 있는 방안들을 담고 있다.

### 2.2 Elasticsearch

Elasticsearch[7]는 본래 Apache Lucene기반의 전문(full-text) 검색엔진이지만, Elasticsearch가 설치된 각각의 노드들을 한데 묶어 구성한 클러스터를 단순 검색엔진의 역할을 넘어서 분산 데이터베이스로도 사용할 수 있다. 분산 데이터베이스로서의 Elasticsearch는 다음과 같은 장점을 가진다. 첫째, 고정 스키마가 없는 JSON문서 형태로 데이터를 관리하기 때문에 데이터의 유연성이 뛰어나다. 둘째, 복잡한 환경설정 없이 외부 시스템 리소스를 추가하여 선형적으로 성능을 증가시킬 수 있기 때문에 확장성이

뛰어나다. 셋째, 실시간성을 가지고 있어 데이터의 저장과 분석을 실시간으로 처리할 수 있다. 마지막으로 원본과 사본의 부분파일인 샤드를 클러스터 내의 여러 노드에 분산 중복 저장함으로써 데이터의 손실을 방지할 수 있다 [9].

### 2.3 콘텐츠 이용 로그 수집·분석 시스템과 프로토타입

그림 1에서 제시한 바와 같이 [4]의 서비스 모델을 구성하는 주체들 가운데는 디지털 콘텐츠의 유통 및 이용내역을 실시간으로 수집 할 수 있는 모니터링 시스템이 존재한다. 이는 디지털 콘텐츠 거래의 투명성과 공정성의 확보를 위한 것으로서 본 연구팀에서는 [4]의 후속연구로 이 모니터링시스템을 설계하기 위한 목적으로 디지털 콘텐츠의 거래상황 모니터링을 위한 로그 수집·분석 시스템을 설계 하였다 [5].



(그림 2) 프로토타입 로그 수집·분석 시스템의 흐름도

[5]의 연구에서 설계한 시스템과 데이터 저장에 대한 조건들을 바탕으로 그림 2와 같은 흐름도를 가지는 프로토타입을 구현하였다 [6]. 프로토타입은 이용자로부터 발생하는 로그를 수집하여 도착한 순서대로 저장하고, 저장한 로그 데이터를 분석 및 시각화 할 수 있도록 구성하였다.

[6]에서 로그 데이터 저장소로는 2.2절에서 언급한 것처럼 유연성과 확장성이 뛰어나고, 실시간 분석을 지원하면서 데이터의 손실을 방지할 수 있는데다 검색엔진으로서의 역할까지 수행할 수 있는 Elasticsearch를 사용하였다.

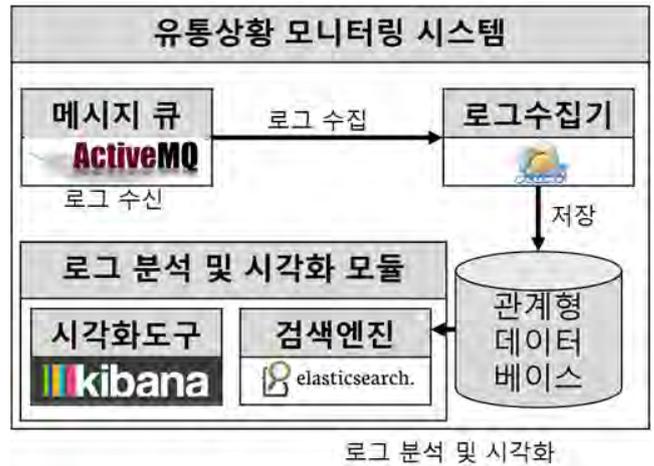
그러나 프로토타입 구현 이후 추가 연구를 통해 Elasticsearch는 데이터 저장소로서의 유연성, 확장성, 실시간성도 뛰어나고 전문 검색엔진으로서의 성능도 뛰어나나 한 시스템에서 Elasticsearch가 저장소와 검색엔진의 두 가지 역할을 동시에 수행하는 경우, 검색엔진으로서의 역할 수행 중 클러스터 내의 노드들이 서로 충돌을 일으키거나 노드가 멈추게 되면 검색엔진으로서의 역할도 어려울 뿐 더러 저장하고 있던 데이터들의 손실까지 발생하여 데이터 저장소로서의 역할도 할 수 없음을 발견하였다 [8]. 따라서 Elasticsearch를 검색엔진과 로그저장소로 동시에 사용한 [6]의 설계상으로는 저장하고 있는 로그 데이터의 안정성을 보장할 수 없었다. 따라서 [6]의 설계를 기반으로 저장하고 있는 로그의 안정성을 확보할 수 있는 새로운 로그 저장소의 설계 및 데이터안정성 증진방안이 필요하다.

### 3. 로그 저장 장치의 안정성 증진 방안

본 논문에서는 수집한 로그데이터의 안정성을 확보하기 위해 프로토타입에서 저장장치와 검색엔진으로 사용하던 Elasticsearch의 역할을 축소하고 로그 데이터를 관계형 데이터베이스 관리 시스템에 저장하는 디지털 콘텐츠 거래 모니터링 시스템의 로그 저장 장치를 제안한다. [5]의 설계의 구성요소 중 하나인 로그 데이터 저장소는 저장하고 있는 데이터의 안정성을 필히 보장해 주어야 하는데 트랜잭션의 원자성, 일관성, 고립성, 영구성(ACID : Atomicity, Consistency, Isolation, Durability)을 잘 지원하는 시스템이 관계형 데이터베이스 관리 시스템이기 때문에 이를 사용하고자 한다.

그림 2에서 표시된 바와 같이 프로토타입에서는 수집·분석 시스템의 검색엔진과 로그 저장소의 목적으로 Elasticsearch를 사용하였다. Elasticsearch는 실시간으로 로그데이터를 수집하고 분석 할 수 있는 장점이 있다. 그러나 Elasticsearch는 같은 클러스터 내 노드들 간의 충돌이 발생하거나 실행중인 노드가 멈추는 경우 데이터의 안정성을 보장할 수 없다. 따라서 시스템 내에서 Elasticsearch의 역할을 로그 데이터의 분석과 빠른 확인을 위한 검색엔진으로만 동작하도록 그림 3과 같이 그 역할을 축소하였다. 이를 통해 검색 및 분석 과정 중 Elasticsearch가 장애를 일으키더라도 로그 데이터는 영향을 받지 않도록 설계하였다.

앞서 언급하였던 Elasticsearch가 저장 장치로서 보여주었던 단점인 데이터 손실을 방지하기 위해 본 연구에서는



(그림 3) 개선된 수집·분석 시스템의 흐름도

로그 데이터 저장 장치로 관계형 데이터베이스 관리 시스템을 사용하였다. 관계형 데이터베이스 관리 시스템은 트랜잭션의 ACID 특성을 엄격하게 지원하기 때문에 저장하고 있는 로그 데이터의 안정성을 보장할 수 있다. 또한 관계형 In-Memory 데이터베이스를 채택하면 빠른 속도로 로그데이터의 입출력을 처리할 수 있어 데이터의 안정성 및 실시간 수집 및 분석을 지원해야 하는 로그 저장 장치의 조건에 부합한다.

제안한 로그 저장소에 저장하는 콘텐츠 이용 로그를 위한 스키마의 예는 그림 4와 같다. 그림 4의 로그 스키마는 콘텐츠를 구분하기 위한 UCI 코드, 콘텐츠 공급자를 확인하기 위한 OSP, 콘텐츠 이용자 정보를 담고 있는 이용자, 콘텐츠 이용 일시를 확인하기 위한 이용일시, 이용자가 어떤 형태로 콘텐츠를 사용했는지 나타내는 제공 형태, 그리고 추가설명을 담고 있는 기타로 이루어져 있다. 모니터링 시스템은 그림 4의 스키마를 적용한 로그들을 통해 콘텐츠별 이용현황, OSP별 콘텐츠 공급 현황, 이용자별 사용 현황 등을 분석할 수 있는 자료를 제공한다.

UCI	OSP	이용자	이용일시	제공형태	기타
-----	-----	-----	------	------	----

(그림 4) 콘텐츠 이용로그 스키마의 예시

트랜잭션이 ACID를 엄격하게 만족해 데이터의 안정성을 보장하기 때문에 관계형 데이터베이스는 성능 향상을 위한 확장성을 제공하기 어려워 다수의 이용자들이 보내 온 로그를 바로 저장하기 어렵다는 단점이 있다. 하지만 프로토타입에서 다수의 이용자들로부터 수집해 오는 로그를 안전하게 수집하기 위해 사용한 메시지 큐와 로그수집기를 사용하면, 제안한 로그 저장 장치도 불특정 다수에게서 도달하는 로그를 안전하게 수집하면서 로그 저장을 위

해 필요한 데이터베이스의 트랜잭션을 줄여 성능을 향상시킬 수 있다.

그림 3에 나타난 봐와 같이 제안된 로그 저장 장치의 관계형 데이터베이스는 로그 데이터 수집을 위한 메시지 큐, 그리고 수집한 로그를 저장하기 위한 로그 수집기와 연결되어 있다. 메시지 큐는 외부에서 오는 로그데이터의 수집을 위해 이용자가 전송한 로그데이터들이 시스템에 도착하는 순서대로 로그데이터를 로그 수집기로 보낸다. 로그 수집기는 메시지 큐로부터 로그데이터를 받아 보관하고 있는 로그 데이터의 수 혹은 로그 데이터의 대기시간이 특정한 조건을 만족하면 저장 장치에 로그 데이터를 저장한다. 단 로그 수집기가 임시로 로그를 보관하고 있는 시간은 실시간 로그 수집 및 분석에 지장이 없어야 한다. 이처럼 메시지 큐와 로그 수집기는 저장장치인 관계형 데이터베이스에서 발생하는 트랜잭션의 수를 적절하게 유지시켜 관계형 데이터베이스에서 트랜잭션의 수행시간에 의한 병목현상을 줄여 줌으로써 확장성이 부족한 관계형 데이터베이스의 단점을 개선하면서 다수의 이용자들로부터 로그를 수집할 수 있다.

#### 4. 결론 및 향후연구

본 논문에서는 Elasticsearch 대신 관계형 데이터베이스를 적용한 콘텐츠 거래 모니터링을 위한 로그 수집·분석 시스템의 로그 저장 장치를 제안하였다. 제안한 로그 저장 장치는 프로토타입의 로그 저장 장치와는 달리 데이터의 안정성을 보장하면서 검색엔진이 동작 중 멈추더라도 로그를 저장 할 수 있다. 또한 관계형 데이터베이스가 수행해야 하는 트랜잭션의 양을 조절함으로써 모니터링 시스템이 다수의 이용자들로부터 수집하는 로그 데이터들을 보다 안정적으로 처리할 수 있다.

향후연구로 콘텐츠 거래 모니터링을 위한 로그 수집·분석 시스템에서 실시간으로 로그를 수집하고 분석할 수 있게 하는 로그 수집기의 설정 값을 정하는 실험을 진행 할 예정이다. 또한 실시간으로 수집되는 데이터의 양에 따라 관계형 데이터베이스가 효율적으로 데이터를 저장, 처리할 수 있는 확장성을 강화하는 연구를 진행 할 예정이다.

#### 참고문헌

[1] 한국콘텐츠진흥원 “2014 한국콘텐츠 산업 백서” 한국 콘텐츠 진흥원, 2015  
 [2] 박홍근, 웹하드, 100건 중 54건 정산누락 불법 부당이득, 2013 국정감사 보도자료  
 [3] 노승욱 “iMBC와 웹하드 업체들 간 저작권료 갈등…다 운로드 정산 누락 둘러싸고 신경전, 매경 이코노미”

2014[Online] <http://news.mk.co.kr/v2/economy/view.php?sc=50000006&cm=%C4%AE%B7%B3&year=2014&no=212397&relatedcode=&wonNo=212975>  
 [4] 변은정, 송광호, 김유성, “분배정산의 투명성 재고를 위한 국가표준식별체계 기반 디지털 콘텐츠 유통 서비스 모델” 제 41회 한국정보과학회 동계학술대회 논문집. 2014 41(2). pp.1266-1268  
 [5] 변은정, 송광호, 김유성 “디지털 콘텐츠의 유통상황 모니터링을 위한 로그수집 분석 시스템 설계” 2015년도 한국빅데이터서비스학회 동계학술대회 발표논문집. pp.14-17  
 [6] Kwangho Song, Jihong Min, and Yoo-Sung Kim, “Development of a Web-based Digital Content Distribution Monitoring System Using Universal Content Identifier” Proceeding of International Symposium on Engineering and Applied Science. 2015. pp. 186-192  
 [7] <https://www.elastic.co/products/elasticsearch>  
 [8] Aphyr, “Call me maybe:Elasticsearch 1.5.0” 2015 [Online]<https://aphyr.com/posts/323-call-me-maybe-elasticsearch-1-5-0>  
 [9] Oleksii Kononenko, Olga Baysal, Reid Holmes, and Michael W. Godfrey “Mining Modern Repositories with Elasticsearch” Proceeding of the 11<sup>th</sup> Working Conference on Mining Software Repositories. 2015. pp. 328-331