디지털저작물 저작권보호시스템

A Study on Systems to Protect Copyright of Digital Contents

김 용(Yong Kim)*
남 궁 황(Hwang NamKoong)**

목 차

초 록

디지털기술의 급속한 발전으로 인해 현재 대부분의 문서가 디지털화된 형태를 이루고 있으며 이러한 정보는 지속적으로 증가하고 있는 상황이다. 웹 및 문서저작도구의 발전과 함께, 정보의 생성과 공유가 쉬워지면서 중복적으로 존재하는 정보의 비율이 갈수록 높아지고 있으며 일부를 표절하여 자신의 정보로 사용하는 문서의 불법적인 복제문제가 발생할 수 있다. 현재 온라인 상에서 제공되고 있는 수 많은 정보는 그것을 접하는 사용자들에게 유용하게 사용될 수도 있지만 불법복제(illegal copy)나 표절(plagiarism)과 같은 형태로 이용될 수 있는 가능성도 높다. 두 경우에 있어서 원문의 부분 또는 전체를 그대로 사용하는 경우가 있으며, 특히, 표절의 경우에 있어서는 문장의 재구성, 유사단어로 대체하는 것처럼 원문과는 다른 형태로 나타날 수 있다. 그러나 표절에 있어서 엄청난 양의 정보중에서 일부를 표절한 사실을 알아내기란 쉽지가 않다. 왜냐하면 표절을 판별하기 위해서는 기존에 존재하는 모든 정보를 알고 있어야 하는데 이것은 이론상으로 사람의 힘으로는 불가능하기 때문이다. 또한 저작자의 동의 없이 이루어지는 불법적인 복제는 디지털콘텐츠의 유통을 위한 커다란 걸림돌이 되고 있다. 따라서 기존의 문서와의 유사성 판별을 통해서 자동적으로 표절의 가능성을 제시해 줄 수 있는 기술과 함께 근본적으로 디지털 저작물에 대한 불법적인 복제를 막을 수 있는 방법이 필요하다.

키워드: Digital Right Management, copy detection, watermarking, 저작권, 전자도서관

* 한국통신 멀티미디어연구소 EOD연구실
** 국방부 합동참모본부
접수일자 2000년 12월 6일

1. 서론

컴퓨터 기술의 혁신적인 발전과 인터넷 등의 새로운 정보통신기술은 정보화 사회로의 새로운 전기를 마련할 수 있는 기반 여건을 제공하고 있다. 이러한 제반여건과 함께 디지털기술의 급속한 발전으로 인해 현재 생산되는 대부분의 문서가 디지털화된 형태를 이루고 있으며 이러한 정보는 기하급수적으로 증가하고 있는 상황이다. 또한, 전자출판, 컴퓨터, 통신, 멀티미디어 기술의 발전으로 문서, 음성, 사진 및 비디오 데이터 등 다양한 매체들은 전자기적 장치에 의하여 디지털화 되어 효율적으로 저장, 접근, 이용이 가능하게 되었고, 많은 양의 정보를 디지털 형식으로 저장·전송하도록 허용하였다. 따라서 데이터의 형태가 점차로 아날로그 형태에서 디지털 형태로 변하고 있으며, 이러한 추세는 World Wide Web의 출현으로 더욱 증가되었다. 디지털형식은 정보의 저장이나 변환이 편리하기 때문에 가상공간에서의 지적재산보호에 어려움이 있다. 그리고 데이터의 디지털화와 멀티미디어의 발달, 인터넷의 보급으로 인한 전자상거래와 같은 가상의 상거래시장이 주목을 받으면서 디지털 데이터의 복제가 확산됨에 따라 여러 가지 멀티미디어 데이터에 대한 소유권 문제와 이를 효율적으로 보호할 수 있는 기술이 요구되고 있다. 그러나 현재 디지털 영상물의 저작권 보호(copyright protection)와 인증(authentication)에 대한 해결책은 아직도 인정할 만한 방법이 제시되지 않고 있는 실정이다. 이러한 현실에서, 폭발적인 디지털정보의 증가는 정보의 신속하고 정확한 제공을 목적으로 하는 도서관이나 기존의 웹상의 정보검색시스템의 한계를 뛰어넘을 정도로 증가하고 있다.

1.1 연구 목적

현재 다양한 멀티미디어 정보를 포함하고 있는 전자문헌들을 포함한 멀티미디어 자원을 인터넷상에서 우리는 자주 접할 수 있다. 이러한 디지털저작물의 폭발적인 증가와 함께, 이에 대한 저작권보호라는 측면이 주요한 관심사로 떠오르고 있다. 이전의 저작권 보호라는 관점이 주로 물리적인 도서나 상품등에 그 초점이 맞추어 졌으며 이에 대한 보완책으로서 주로 법적, 제도적인 측면에서 많은 노력들이 이루어진 반면, 근래에 들어서는 위에서도 언급한 바와 같이 정보기술의 발전과 WWW의 급속한 확산으로 인하여 과거의 저작물의 형식과는 다른 디지털저작물의 확산과 함께 이러한 디지털저작물을 보호할 수 있는 기술적인 분야가 새로운 관심분야로 부각되고 있다. 〈표 1〉은 콘텐츠를 제작하는데 있어서 전통적인 방식과 디지털방식의 차이점을 보여주고 있다. 특히, 정보와 지식이 중요한 경제적가치로서 평가되어지고 있는 시점에서 이러한 지식과 정보의 저장소로서 전자 도서관의 중요한 기능을 보다 효율적으로 수행하고 발전 시키기 위해서는 정보의 유통 및 공유라는 기본 목적을 충족하기 위한 중요한 요소 중의 하나로서 디지털컨텐츠에 대한 효과적인 저작권 보호장치를 갖추어야 한다는 것이다. 만일 이러한 보호장치를 갖추지 못한다면 저작권 소유자들은 온라인을 통한 콘텐츠 공급과 새로운 미디어를 이용한 콘텐츠의 생산을 꺼리게 될 것이다. 현재 온라인 상에서 제공되고 있는 수 많은 정보는 그것을 접하는 사용자들에게 유용하게

사용될 수 도 있지만 표절(plagiarism) 또는 복제(illegal copy)와 같은 불법적인 형태로 이용될 수 있는 가능성도 높다. 표절은 원문의 부분 또는 전체를 그대로 사용하는 경우가 있으며, 문장의 재구성, 유사단어로 대체하는 것처럼 원문과는 다른 형태로 나타날 수 있다. 그러나 표절의 방법과는 상관없이 엄청난 양의 정보 중에서 일부를 표절한 사실을 알아내기란 쉽지가 않다. 왜냐하면 표절을 판별하기 위해서는 기존에 존재하는 모든 정보를 알고 있어야 하는데 이것은 이론상으로 사람의 힘으로는 불가능하기 때문이다. 따라서 기존의 문서와의 유사성 판별을 통해서 자동적으로 표절의 가능성을 제시해 줄 수 있는 기술과 함께 근본적으로 디지털 저작물에 대한 불법적인 복제를 막을 수 있는 해법이 필요하다. 한편, 세계적으로 많은 국가들은 국가기반 정보화사업의 일환으로 각종 정보·저작물들의 디지털화 및 서비스를 적극 추진 중이며 이를 추가 지원하려는 계획을 세우고 있다. 현재 국내에서 진행중인 국회와 중앙도서관의 소장자료에 대한 디지털화가 대표적인 사례라 할 것이다. 그러나 정부는 저작물의 디지털화 및 서비스를 가속화함에 있어 정작 문제의 핵심인 저작권 처리에 대해서는 뚜렷한 해결책을 제시하고 있지 못하고 있다. 현저작권법은 도서관에게 저작자의 허락 없이 저작물을 디지털화할 수 있는 권리를 부여하고 있지 않다. 각종 멀티미디어 저작물을 제작하기 위해서는 수많은 기존 저작물의 이용이 필수적이며, 이러한 기존 저작물에 대한 수요는 정보문화산업의 발전과 함께 급증하고 있다. 그러나 현재 증가하는 저작물 수요 요구에 대해 시의적절하게, 효과적으로 공급해 줄 수 있는 창구가 마련되고 있지 않으며, 이러한 상황은 필수불가결한 저작권 침해를 양산하고 있다. 예를 들어 인터넷상에서 가장 많은 복제가 일어나고 있는 MP3화일의 경우에도 원저작자의 동의 없이 공공연하게 불법적인 복제가 일어나고 있는 것도 주지의 사실이다.

디지털 정보의 보호를 위해 적용할 수 있는 방법은 크게 다음 세 가지로 분류할 수 있다.

첫째, 기존의 공통키 또는 공개키 암호화 알고리즘을 이용하여 주어진 데이터를 암호화하는 방법으로, 영상을 원래의 데이터로 복구하기 위해서는 관련 키를 알고 있어야 한다. 이 방법은 수학적으로는 안전하나 사람이 개인키로 암호화된 정보를 배포하는 것을 막을 수 없다는 단점을 가지고 있다. 그리고 단순한 공통키나 공개키 암호화만으로는 저작권을 위반하는 사항을 완벽하게 막을 수 없다.

〈표 1〉 콘텐츠 제작을 위한 기술적 특징

구분	전통적 기술	디지털 기술
복제의 용이성	노력과 시간 소요	용이, 신속
복제의 질	질의 저하	원본과 동일
조작과 변경	어렵고 흔적이 남음	쉽고 흔적이 남지 않음
저작물의 융합	종류, 매체별로 존재	멀티미디어 저작물
대중에의 전달	시간과 공간의 제약	시간, 공간의 제약이 없음

둘째, 보호 대상 영상정보에 대하여 접근제어용 방화벽(firewall)을 구축하는 방법으로, 컴퓨터 네트워크를 통한 사용자 인증 절차를 거쳐 영상 데이터의 사용을 제한하는 방법이다. 이 방법 역시 사용자가 임의로 영상 자료를 배포하는 것을 막을 수가 없다는 단점을 가지고 있다.

셋째, 디지털 영상의 불법적인 내용 조작을 막고, 영상의 소유권을 보장할 수 있는 방법으로 디지털 워터마크(digital watermark)가 있다. 디지털 워터마크는 공개키 알고리즘이나 방화벽 등으로 해독된 영상에 대하여 부가적인 보호를 제공한다. 저작권 정보, 배포자 정보 그리고 사용자 정보를 영상에 삽입함으로써 법적인 문제가 발생하였을 때 해결책을 제시할 수 있다. 그러나 워터마크 기술에 대해서는 아직 많은 연구가 필요한 실정이다.

따라서 본 연구에서는 이러한 문제점을 인식하고 이를 해결하기 위한 기술적분야와 이러한 보호기술에 대한 적응분야에 대하여 알아보고자 한다.

2. 디지털저작물 저작권보호기술

2.1 연구동향

멀티미디어 콘텐츠 저작권 보호와 관련된 기반기술 연구로서는 1990년 중반부터 시작되어 최근 디지털저작물의 폭발적인 증가에 따른 저작권 보호에 대한 관심이 증가와 함께 중요한 연구분야로서 관심을 모으고 있다. 디지털저작물에 대한 표절 혹은 불법복제 확인 관련 시스템은 크게 선방지(prevention)시스템과 후판별(detection)시스템으로 구분할 수 있다. 방지시스템은 권한이 없는 사람으로부터의 표절이나 복제를 사전에 예방하는 것으로 암호키를 이용하는 사용자인증 또는 콘텐츠에 대한 암호화 등의 기법이 대표적으로 이용되고 있으며 후판별시스템으로서 멀티미디어콘텐츠가 주대상인 디지털 워터마크(digital watermarking: 전자은화, 전자투과)는 국내외적으로 많은 연구가 진행중에 있으며 이와 더불어 텍스트콘텐츠를 주대상으로 하고 있는 복제판별방법(copy detection)은 불법으로 복제된 문서의 유통과정에서 이를 판별하여 문서의 유통을 방지하는 시스템으로서 최근에 스탠포드대학을 중심으로 연구가 진행 중에 있다. 이러한 요소기술들을 바탕으로 전자상거래시스템에서 이용되고 있는 사용자인증 등의 방법과 함께 디지털저작물의 유통 및 관리에 대한 포괄적인 해결점을 찾고자 하는 디지털저작권관리시스템(digital right management) 분야가 있다.

이러한 관점에서 본 연구에서는 인간의 지적 창작 활동으로 생산된 다양한 멀티미디어 콘텐츠(이미지, 오디오, 비디오, 영화, 음악, 텍스트 등)의 복제를 근본적으로 억제하고 유통을 제한하여, 지적 창작물의 저작권을 보호할 수 있는 효과적인 방법에 대한 연구동향과 함께 그 기대효과에 대하여 알아보고자 한다. DRM(Digital Rights Management) 시스템은 디지털 콘텐츠의 주문, 인증, 결제를 일괄적으로 제공하는 저작권 관리 솔루션으로서 디지털콘텐츠의 유통에 대한 모든 분야에 걸쳐 다양한 보호기술들이 적용되어지며 콘텐츠에 대한 암호화 및 복호화(encryption/decryption), 사용자인증(authentication), 키 교환/관리, 워터마킹(watermarking)과 핑거프린팅(fingerprinting) 등의 다양한 기술들이 적용되어진다.

디지털워터마크 기술은 DRM을 구성하는 주요 기술로서 콘텐츠의 불법 복제를 방지하는 방법으로서 주로 이미지/비디오/오디오/텍스트(source code) 등의 멀티미디어 저작물에 적용이 가능하므로서 보호기술분야 중에서 가장 활발하게 연구가 진행이 되고 있는 분야이다. 한편 주로 텍스트기반의 문서에 대한 복제를 판별하는 방법으로서 텍스트용 문서복제 판별기술(copy detection)이 있는데 디지털워터마크 기술이 일반적인 텍스트 저작물에 적용이 가능하지만 개별적인 모든 글자에 워터마크를 삽입한다는 것이 현실적으로 불가능하다. 따라서 문서의 구조적, 논리적인 측면을 고려하여 원문서와의 유사도를 비교하여 문서의 복제여부를 판별하는 방법으로서 문서복제 판별기술(Copy Detection)분야가 텍스트용 문서의 저작권을 보호하기 위한 방법으로서 제시되고 있다. 각 분야별 구체적인 특징은 다음과 같다.

2.2 DRM(Digital Right Management)

2.2.1 개 요

문서의 불법 복제를 불가능하게 하고 디지털 저작물의 유통 및 관리에 대한 해결방법을 제공하기 위한기술로서 디지털저작물의 인터넷을 통한 유통 및 대금지불 방법들을 포함하고 있다. 콘텐츠에 대한 암호화 및 복호화(encryption/decryption)기능을 수행하며, 사용자인증(authentication), 콘텐츠 암호화와 복호화 및 사용자 인증등에 관여하는 공개키 및 비밀키의 교환 및 관리기능을 포함하고 있으며, 실질적인 콘텐츠에 대한 디지털워터마킹과 핑거프린팅(watermarking and fingerprinting) 등의 다양한 기술을 이용하여 이를 삽입하므로서 디지털저작물의 저작권의 보호를 위해 필요한 기능을 수행할 수 있는 저작권보호시스템이라고 할 수 있다.

2.2.2 기 능

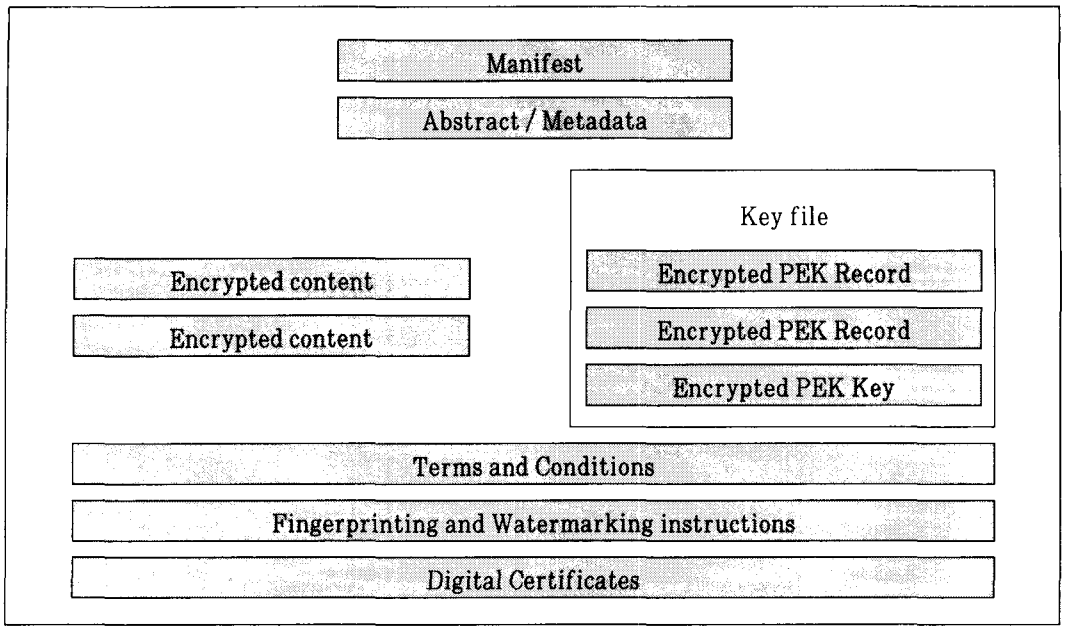
DRM 시스템은 디지털콘텐츠의 포괄적인 저작권보호를 위한 요소기술들을 적용하므로서 통합된 해법을 제시하고 있다. 이러한 DRM시스템이 효율적인 저작권보호를 위하여 제공하고 있는 기능은 아래와 같다.

- 중요한 디지털콘텐츠를 제공하기 위한 암호키 인증 관리 솔루션을 적용하므로서 불법복제와 표절로부터 저작물을 보호하기 위한 해법을 제공한다.
- 디지털콘텐츠의 유통활성화를 위한 지불관리솔루션을 제공하므로서 디지털콘텐츠의 유료화를 위한 기반을 제공하고 있다.
- 불법적인 복제와 표절에 대한 선방지적인 성격의 암호키 인증관리솔루션과 후판별적인 성격의 워터마킹기술과 함께 강력한 불법복제방지를 위한 시너지효과를 거둘 수 있다.
- 워터마크와 DOI 시스템을 적용하므로서 불법복제물의 효과적인 단속이 가능하다.

2.2.3 시스템 구성

전체적인 시스템의 구성은 각 요소별로 구분이 될 수 있으며 〈그림 1〉은 각 모듈별 관계를 보여주고 있다. 각 모듈별 기능을 알아보면 다음과 같다.

- Manifest: 시스템 내부의 모든 요소들을 암호화된 checksum과 함께 나열한다.



〈그림 1〉 일반적인 DRM시스템의 구조

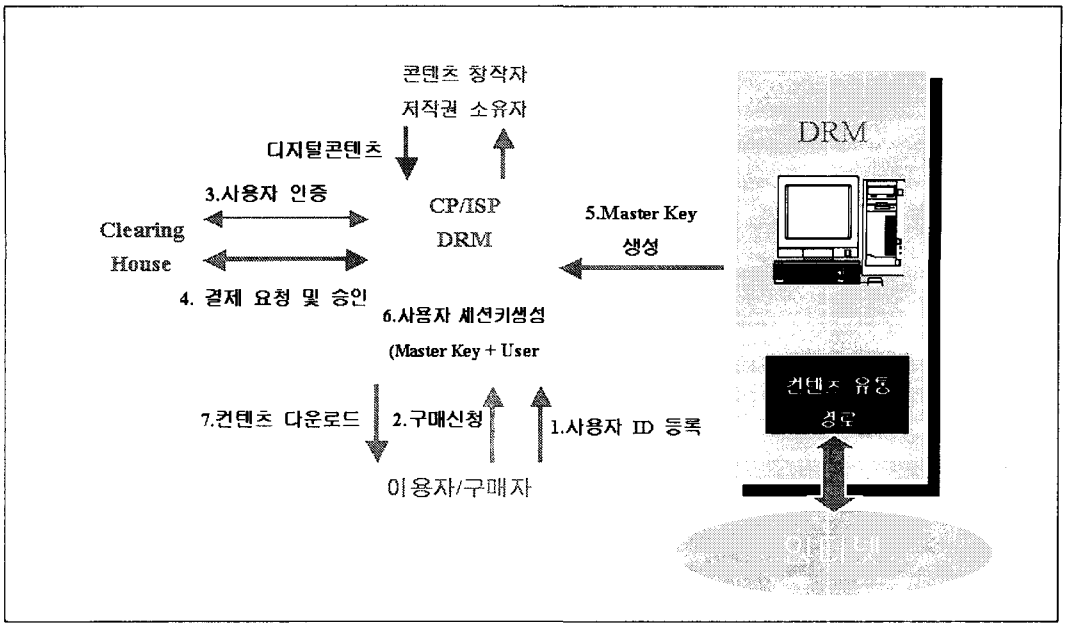
- 초록(Abstract): 고객의 구매의사 결정을 돕기 위해 암호화된 내용의 요약문을 평문의 형태로 제공함
- 메타데이터(Metadata): 문서 내용에 대한 색인 정보를 제공(저자, file size, ...)
- 암호화된 콘텐츠(Encrypted content): 암호화된 문서의 실제 내용물을 담음
- PEK(part encryption key): 문서별로 생성된 암호키로서 마스터키에 의해 암호화되었고 키관리화일내에 저장되어 있음
- Terms and Conditions: 문서에 대한 지적재산권을 기술하고 있다
- Fingerprinting and watermarking: 문서의 원본을 표시하거나 복사본의 경로를 표시하기 위한 digital marking이다
- 디지털인증서: 문서의 내용과 사용자를 인증하는 인증서

한편 〈그림 2〉는 이러한 DRM시스템을 활용하여 디지털 콘텐츠의 유통과정을 기술한 것으로서 이용자가 콘텐츠를 입수하기 위해서는 이용자는 먼저 사용자 등록을 한 후 적절한 사용자 인증단계를 거쳐 DRM을 통하여 콘텐츠에 대한 복호화를 위한 사용자 세션키를 생성한 후 콘텐츠를 입수하게 된다. 이러한 일련의 과정은 전자상거래에서 물품을 구입하는 과정에서의 사용자인증 과정과 동일하다고 할 수 있으며 다른 점은 콘텐츠에 워터마크 또는 핑거프린트 등의 디지털콘텐츠의 복제를 방지할 수 있는 요소가 삽입된다.

2.3 문서복제판별기법(Copy Detection)

2.3.1 개 요

문서복제판별기술은 전통적으로 정보검색기



〈그림 2〉 DRM을 통한 콘텐츠 유통 모델

법에서 사용되는 다양한 기술들을 이용하여 문서간의 유사도를 측정하여 설정된 기준치와의 비교를 통하여 유사성을 파악하는 방법으로서 통계정보를 이용하는 분류기법을 활용할 수 있다. 일반적으로 워터마킹기법이 디지털콘텐츠에 저작자가 직접 특정한 코드나 기호를 삽입하여 이를 판별하므로서 복제를 방지하는 반면에 문서복제판별기법은 기존에 존재하는 다양한 디지털콘텐츠를 다른 특정한 노력 없이 단지 문서간의 유사도를 측정하여 이를 판별하므로서 저작자의 부가적인 노력이 필요 없이 문서복제를 판별하므로서 디지털콘텐츠의 저작권을 보호할 수 있는 기능을 수행할 수 있다. 또한, 문서복제판별기술은 인터넷상의 폭발적인 정보의 증가와 정보의 중복으로 인해 전자도서관 뿐만이 아니라 인터넷상의 정보검색시스템의 효율에 많은 역효과를 나타내고 있는 정보중복의 문제를 해결 할 수 있다. 즉 온라인상에서 존재하거나 유통되고 있는 정보중에서 많은 내용이 중복 저장되고 있기 때문에 정보검색엔진이 동일하거나 유사한 정보를 사용자에게 제공하게 되는 경우가 많다. 이는 정보검색시스템의 성능을 저하시키는 원인이 되며 사용자에게는 같은 내용을 탐색하게 하므로서 불필요한 시간낭비를 초래하게 된다. 따라서 문서검색과정에서 내용적, 구조적 중복성을 자동적으로 판단하므로서 사용자에게 유일한 정보만을 제시할 필요성이 있다.

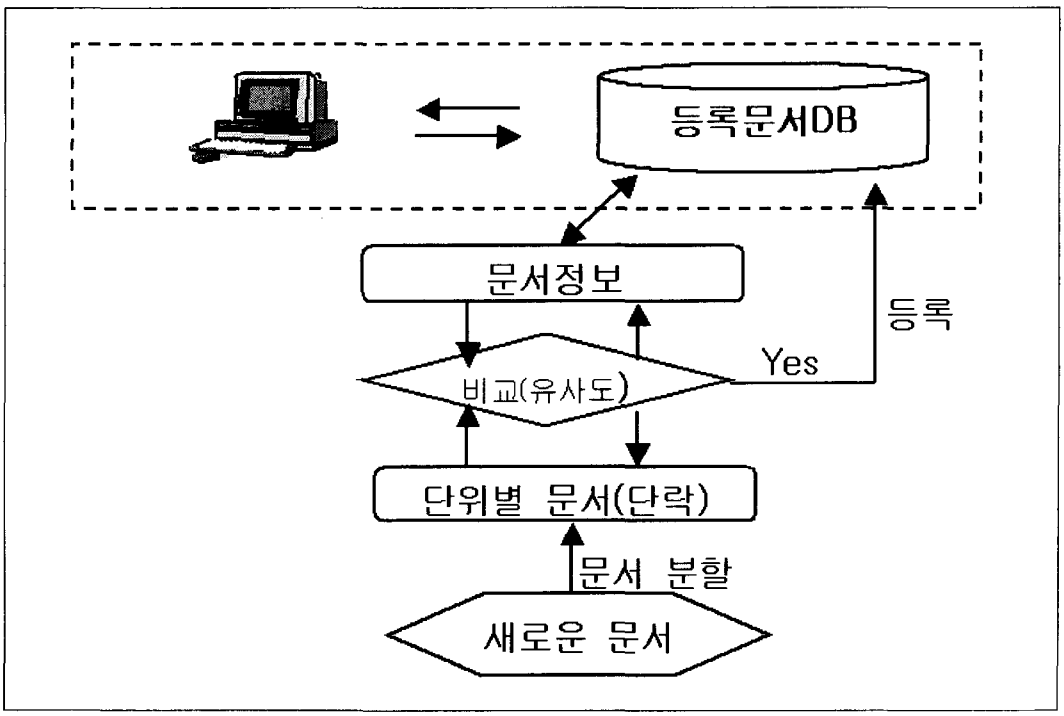
이러한 문서복제판별시스템의 구현 사례로서 스탠포드대학은 미국DARPA의 지원으로 COPS(Copy Protection System)와 SCAM (Stanford Analysis Mechanism) 두개의 시스템을 개발하였다. COPS는 문서등록서버를 구축한 후 등록된 문서에 대하여 문장단위로 해쉬테이블을 생성한다. 새로운 문서가 등록될 경

우 서버 문서와 마찬가지로 해쉬테이블을 만들어 서버 내의 해쉬값과 비교하여 특정 임계값이상이면 중복문서로 판별하는 알고리즘을 가지고 있다. 이러한 표절판별기법과 유사한 연구로서는 HTML문서를 대상으로 DARPA의 지원으로 UCLA에서 수행되어진 연구가 있다. 이 연구는 문서내용의 지역성과 구조를 기반으로 분할된 부분의 유사성을 측정하는 기법을 제시하고 있다.

2.3.2 문서복제판별시스템

〈그림 3〉은 일반적인 문서복제를 판별하는 과정을 보여주고 있는데 먼저 이용자가 문서를 등록하면 시스템은 해당 문서에 대한 문서등록번호와 이에 대한 색인어를 작성한다. 해당 문서는 구조정보와 함께 문단 또는 단락 등을 단위로 나뉘어지며 각 문서단위(chunk)에 대한 등록정보가 부여가 되며 등록정보는 실제 문서정보와 연결되어져 있다. 이러한 과정을 통하여 저장소에 문서와 해당 문서정보가 저장이 된다. 한편 새로운 문서에 대한 등록요청이 들어오면 시스템은 먼저 기존에 저장된 문서에 대한 단위와 새로운 문서의 단위에 대한 유사도를 비교하면서 일정값이상의 유사값이 측정되어지는 경우 해당 문서는 복제된 것으로 판정이 된다. 이러한 유사도의 판별은 문서전체적인 유사도 뿐만이 아니라 문서의 각 개별단위에 대한 유사도의 판별이 가능하다.

이러한 문서복제시스템의 대표적인 구현시스템으로서 SCAM(Stanford Copy Analysis



〈그림 3〉 일반적인 복제문서판별흐름도

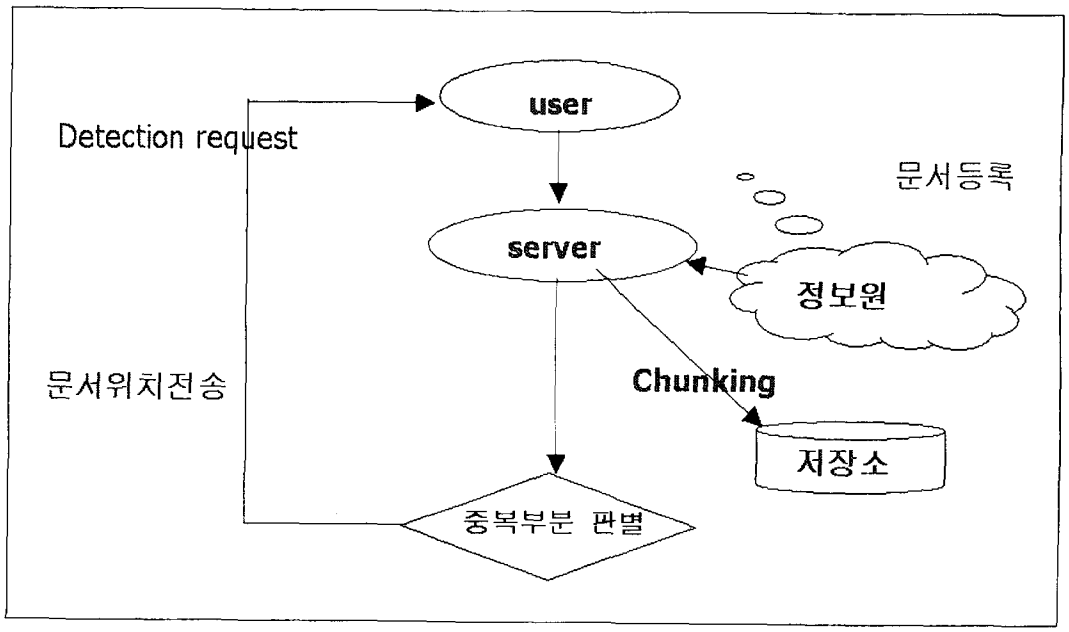
Mechanism)이 있다. SCAM 시스템은 스탠포드 대학에서 대용량의 데이터베이스에 문서를 등록하고 새로이 등록되어지는 문서에 대하여 기존의 문서와의 유사도를 비교하여 문서의 복제여부를 판별하기 위하여 개발된 시스템으로서 주로 대규모의 전자도서관이나 문서저장소와 같은 대용량의 정보저장소에 문서를 등록한 후 새로운 문서가 등록되어지는 경우 기존의 문서와 새로운 문서간의 유사도 비교를 통하여 문서의 복제 여부를 판별할 수 있다. 〈그림 4〉는 SCAM시스템에서 문서에 대한 복제여부를 검출해내는 과정을 보여주고 있다.

SCAM 시스템은 다음과 같은 모듈로 구성이 되어져 있으며 이러한 모듈은 일반적인 문서복제시스템에 기본적으로 제공되어져야 하는 요소 및 기능들이다.

- 버퍼관리자 모듈(Buffer manager): 디스크의 자료를 메인메모리의 버퍼에 저장(caching)하여 관리
- 버퍼풀(Buffer Pool): cache memory
- 역색인화일(Inverted Index File): chunk내의 단어출현빈도(term frequency)와 문서 등록번호를 저장
- Name Mapper: 문서 번호와 실제 문서 위치정보(Web의 URL 등)를 연결해 줌
- Relevance Array: 새로운 문서에 대한 질의어를 생성하고 기존문서에 대한 유사도를 판별하는 과정에서 중복성이 발견될 때 증가되는 유사도를 저장
- 저장소(Repository disk): UNIX file system으로 문서들을 page 형태로 저장

2.3.3 문서복제판별시스템의 특징

- 후판별시스템(post-detection system)



〈그림 4〉 SCAM 시스템의 구조

- 지속적으로 검색이 이루어지므로 언젠가는 색출될 가능성이 높다.
- Internet 상에서 유통되는 불법 복제된 문서들을 색출
- 사용자가 중복의 기준을 정할 수 있다.
- 불법 복제의 판정기준을 절, 문장, 문단 등으로 사용자가 정의할 수 있다.
- 저장되어지는 자료는 문서단위가 아닌 chunk단위
- 구조적인 복제 판별이 가능 (구조검색 기법의 적용)
- 방대한 자료를 다루기에 적합한 data structure를 제공
- 전자도서관이나 Archives와 같은 대규모의 자료를 다루는 시스템에 적절
- 주로 텍스트기반의 자료에 적절

2.4 디지털 워터마크(Digital Watermark)

2.4.1 개요

디지털 워터마크는 멀티미디어 콘텐츠보호를 목적으로 멀티미디어 데이터에 삽입된 디지털 코드나 태그를 의미한다. 즉, 디지털 워터마크는 저작권과 관련된 정보(창작자, 사용자, 사용 가능 횟수, 복제 횟수 등)를 인간의 눈에 쉽게 노출되지 않도록 해당 디지털 미디어에 정교하게 삽입하고, 삽입된 정보를 필요에 따라 추출하여 그 정보를 분석함으로써 저작권자를 확인하는 기술로서, 디지털 워터마크는 인터넷, 인트라넷, 디지털 위성, 디지털 케이블망 등에서 유통되는 이미지, 비디오, 음성, 텍스트 데이터들에 대한 지적 재산권 침해를 방지할 수 있다. 기본적으로, 디지털 워터마크는 크게 저현성(inconspicuousness)과 무작위성(randomness)이라는 두 가지 개념에 근거한다. 모래사장에 섞인 한 톨의 소금은 잘 눈에 띄지 않으며(저현성), 손에서 빠져나간 한 톨의 소금이 모래사장에 떨어진 위치(무작위성)는 짐작하기 어렵다. 전통적 워터마크 개념과 비슷하게, 디지털 워터마크도 멀티미디어 콘텐츠 보호를 목적으로 위의 두 개념에 충실하게 멀티미디어 데이터에 삽입된 코드나 태그를 의미한다. 삽입된 태그는 멀티미디어 데이터의 저작권과 관련된 정보(예를 들면, 제작자, 소유자, 사용자, 미디어 ID, 제작 시기, 복제 가능 횟수, 트랜잭션 ID 등)를 포함하고 있다. 해당 멀티미디어 데이터가 복제될 때 태그도 같이 복제되기 때문에, 저작권에 대한 문제가 발생하였을 때 복제품에 삽입된 태그를 추출하여 정보를 분석함으로써 저작권 시비를 가려낼 수 있다.

2.4.2 디지털워터마크 삽입 및 검출 알고리즘

워터마킹삽입 및 검출알고리즘은 주어진 원영상 I에 레이블(label) S={S1, S2, ・・・,Sn}를, 부호화 과정 E를 통하여 삽입하면 워터마크가 내장된 영상 I′=E(I, S)를 얻을 수 있다. 이때 레이블 S는 영상에 표시된 워터마크가 된다. 테스트 영상 J(워터마크가 삽입되었거나, 삽입되지 않았거나, 손상된 영상)에 대한 소유권을 판정하는 과정은 J 또는 원 영상 I를 입력으로 받아 복호화 과정 D를 통해 레이블 S′=D(I, J)를 추출한다. 추출된 레이블 S′와 S 사이의 유사도를 비교함으로써 소유권을 판정한다. 따라서 부호화 과정, 복호화 과정 그리고 유사도 비교를 어떻게 설정하는가에 따라서 여러 가지의 워터마킹 방법이 존재하게 된다.

2.4.3 디지털워터마크의 요구조건

디지털 저작물에 대하여 워터마크을 입력할 경우 입력되어질 워터마크는 반드시 다음과 같은 기본 요구조건을 만족해야 한다.

- 시각적인 무감지성(Perceptual Invisibility)
  워터마크 내장으로 인한 수정은 시각적으로 이미지의 질을 저하시키지 않아야 한다. 그러나, 시각적으로 차이가 없다 하더라도 원영상을 워터마크된 영상과 비교할 때는 뚜렷해야 한다. 그러므로 원영상은 합법적인 소유자에게만 접근이 가능하고 그런 차이가 관찰자에 의해서는 인식되지 않게 남아 있어야 한다.
- 확실한 추출(Trustworthy Detection)
  워터마크는 어떤 특정한 영상에 대해 충분하고 확실한 소유권 증명을 해야 한다. 워터마크 추출 실패는 나타나지 않아야 하지만, 만약 나타난다면 아주 드물게 나타나야 한다. 워터마크 신호의 특징은 대단한 복잡성을 가진다는 것이다. 이것은 구별이 잘 되는 워터마크들의 광범위한 집합을 만들 수 있도록 하기 위해 필요하다.
- 자동화된 추출과 탐색(Automated Detection/Search)
  워터마크는 한 소유권자의 생산물의 불법적인 파괴에 대하여 네트워크 환경에서 공동으로 접근할 수 있는 영역을 조사하는 탐색 절차와 쉽게 결합되어야 한다.
- 관련된 키(Associated Key)
  워터마크는 "워터마크키(watermark key)"라 불리는 확인 번호와 관계가 있다. 이 키는 워터마크를 만들고, 추출하고, 제거하는데 사용된다. 그 후 이 키는 개인적 소유권의 합법성을 확인하는데 사용된다. 디지털 이미지로부터 추출된 디지털 신호는 워터마크 생성 알고리즘을 통해 어떤 특정 키와 연관되어 있다면 유효한 워터마크라고 가정한다.
- 통계적 무감지성(Statistical Invisibility)
  워터마크는 통계적인 방법을 사용하여 회복되어서는 안된다. 예를 들면 상당히 많이 워터마크된 디지털 영상물들의 소유는 같은 키를 이용하여 통계적인 방법을 적용함으로써 그 워터마크를 배치해서는 안되고 워터마크는 이미지 의존적이어야 한다.
- 다중 워터마킹(Multiple Watermarking)
  같은 이미지에서 다른 워터마크를 충분히 많이 첨가할 수 있어야 한다. 각 워터마크는 유일한 키를 사용하여 추출할 수 있어야 한다. 이런 특징은 이미 워터마크된 영상을 다른 사람이 다시 워터마킹 하는 것을 막을 수 없기 때문에 필요하다. 또 저작권 소유가 한 소유자로부터 다른 소유자로 이동된 경우에 편리하다.
- 강인성(Robustness)
  워터마크가 내장된 디지털 영상은 고의든 고의가 아니든 수정될 수 있다. 따라서 워터마크는 일반적인 영상처리 등의 영상 변형(압축, 여과, 잡음 첨가, 회전, 스케일링 등) 후에도 남아 있어야 한다.

2.4.4 디지털워터마크 생성방법

- 공간 영역에서의 디지털 워터마크
  물리적 픽셀 영역, 즉 공간 영역에서 워터

마크를 삽입하는 가장 간단한 방법은 픽셀들을 임의적으로 선택하여 그것의 밝기 값의 LSB(least significant bit)를 변형시키는 것이다. 이 방법은 잡음과 일반적인 신호처리에 강인(robust)하지 못하다는 단점을 가지고 있다. 또한 데이터 전송 및 잡음(noise)에 매우 민감하고, 데이터 압축과 같은 영상의 변형에 내장된 워터마크를 쉽게 손실하는 문제점이 있다.

이러한 단점을 극복하기 위하여 인간의 시각 특성을 이용할 수 있다. 즉, 인간 시각의 마스킹(masking) 효과에 의해, 영상 내의 결(texture) 영역이나 윤곽선 둘레의 밝기 값의 변화는 육안으로 잘 구별할 수 없다는 점을 이용하여 워터마크를 삽입한다.

- 주파수 영역에서의 디지털 워터마크

영상 데이터를 주파수 공간으로 변환하여 그 주파수 영역들 중에서 시각적으로 덜 민감한 부분에 적응적으로 워터마크를 삽입하는 방법으로서, 단일 주파수 성분을 변화시킴으로써 변환 블록 내의 밝기 값 전체에 영향을 미치고, 따라서 불법적인 공격에 강한 워터마크를 만들 수 있다. 영상 데이터를 주파수 형태로 변형했을 때 가질 수 있는 통신 채널이라고 가정한다면, 워터마크는 그 통신 채널로 통과하는 신호라고 볼 수 있다. 특정 주파수 대역의 에너지는 감지할 수 없을 정도로 작지만 주파수의 위치와 변화량을 알고 있는 소유권자에 의해 산재해 있는 주파수 성분을 모으면 높은 신호대 잡음비(signal to noise ratio)로 신호를 검출할 수 있다. 워터마크를 영상이 갖고 있는 여러 주파수 영역으로 확산시킴으로써 특정 주파수 대역의 에너지는 감지하기 어렵게 한다. 영상의 변화를 감지 못하도록 하면서 시각적으로 중요한 영역에 정보를 삽입하는 알고리즘은, 시각적 변형과 JPEG 압축, 그 외 영상 처리 기술에 대하여 장단점을 고려하여 삽입하고자 하는 주파수 영역을 선택하여 워터마크를 삽입한다.

2.4.5 요소별 기능

- 워터마크 입력기

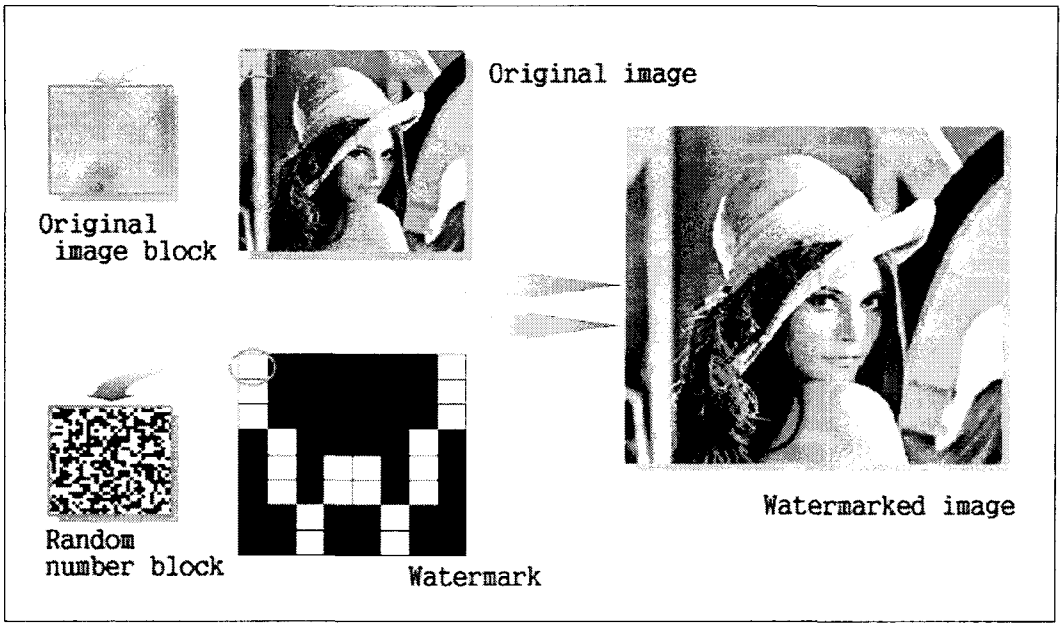
입력기에는 일반 PC나 워크스테이션에서 독립적(stand-alone)으로 작동하거나 이미지 편집 소프트웨어의 플러그인(plug-in)으로 작동 두 가지 방식의 입력기가 사용이 될 수 있으며 이러한 입력기는 다양한 형식의 호스트 데이터에 디지털 워터마크를 입력하는 기능을 수행한다.

- 워터마크 검출기

검출기는 호스트 데이터에서 삽입된 워터마크 정보를 추출하는 디지털 워터마크 검출하는 기능을 수행하는 것으로서 일반 PC나 워크스테이션에서 독립적으로 작동하거나, 이미지 편집 소프트웨어의 플러그인으로 작동하거나 또는 인터넷 브라우저의 플러그인으로 작동하는 크게 세 가지 방식이 있다.

- 워터마크 추적기

인터넷을 탐색하여 워터마크된 멀티미디어 콘텐츠를 주기적으로 추적하는 기능을 수행하는 것으로서 기존의 정보검색엔진이 활용될 수 있다.



〈그림 5〉 워터마크 입력 예

3. 적용분야

디지털저작권 보호기술은 디지털콘텐츠의 증가와 함께 다양한 분야에 적용 가능하다.

전자도서관구축을 위한 디지털저작물을 포함하여 이미지, 비디오, 오디오와 같은 멀티미디어 저작물에 대한 유사성을 측정하므로서 불법 복제된 저작물에 대한 유통을 방지 할 수 있으며 또한 저작물에 저작권 정보를 디지털 코드 형태로 은밀하게 삽입하여 저작권의 진위와 관련된 분쟁을 효과적으로 해결할 수 있다. 또한 정보유통의 활성화 측면에 있어서 멀티미디어 콘텐츠의 저작권이 보호되고 불법으로 복제된 콘텐츠의 추적이 가능하면, 네트워크를 통한 멀티미디어 콘텐츠의 유통을 활성화하는데 크게 기여할 수 있다. 한편, 네트워크를 통한 정보유통에 있어서 사회적으로 많은 문제가 되고 있는 디지털정보보안 측면에서 디지털 워터마크 기술은 디지털 콘텐츠의 저작권 보호뿐만 아니라, 주석은닉(hidden annotation), 데이터 무결성 증명(integrity verification), 비밀통신(invisible communication) 과 같은 다양한 분야에 적용될 수 있다.

특히, 현재 정부에서 추진하고자 하는 DOI 시스템(The Digital Object Identifier System)의 핵심기술로서 요구되는 분야가 문서복제판별기술과 워터마킹(Watermarking) 기술이다. 미국을 비롯한 선진국에서는 이미 1990년대 중반부터 디지털콘텐츠에 대한 효율적인 접근과 저작권 보호에 대한 필요성을 인식하고 이와 관련된 연구 및 각종 사업을 적극적으로 추진하고 있다. 이를 보다 구체적으로 알아보면 다음과 같다.

3.1 디지털콘텐츠보호

멀티미디어 디지털데이터는 인간 생활과 밀접한 관계를 형성하고 있으며, 그 편리성 때문에 인간의 지적 창작물은 점점 디지털 형태로 표현되어 제작되고 있다. 그러나, 디지털 데이터는 쉽고 완전하게 복제될 수 있다는 치명적인 약점을 갖고 있다. 이런 약점 때문에, 멀티미디어 디지털 데이터가 원래의 창작물인지 아니면 복제물인지를 판별해야 하는 경우가 자주 발생하게 되었고, 보다 근본적으로는 복제물의 무분별한 생산을 억제하거나 설사 복제가 가능하다고 할 지라도 이에 대한 유통을 근본적으로 제한 할 수 있는 방법의 개발이 절실히 필요하다.

따라서 디지털콘텐츠의 불법적인 복제와 유통을 방지하므로서 디지털콘텐츠 생산을 활성화시킬 수 있으며 이를 통하여 사용자들에게는 보다 나은 편의성을 제공할 수 있다.

3.2 정보검색을 위한 중복문서의 판별

정보의 중복으로 인해 나타나는 문제는 현재 인터넷에서 서비스를 제공하고 있는 정보검색시스템에서 가장 심각하게 나타나고 있다. 즉 온라인 상에 저장되고 있는 정보중에서 많은 내용이 중복 저장되고 있기 때문에 정보검색엔진이 동일하거나 유사한 정보를 사용자에게 제공하게 되는 경우가 많다. 이는 정보검색시스템의 성능을 저하시키는 원인이 되며 사용자에게는 같은 내용을 탐색하게 하므로서 불필요한 시간낭비를 초래하게 된다. 따라서 문서검색과정에서 내용적, 구조적 중복성을 자동적으로 판단하므로서 사용자에게 유일한 정보만을 제시할 필요성이 있다.

3.3 전자도서관 정보저장(Repository) 서비스를 위한 중복문서의 제거

대량의 온라인 정보를 저장하게 되는 전자도서관의 경우 수시로 새로운 정보를 저장하게 되는데 기존에 저장된 것과 동일한 내용이 중복 저장되는 문제가 발생할 수 있다. 특히, 근래의 전자 도서관구조에서는 정보저장서비스, 탐색서비스, 클라이언트서비스, 등이 독립적으로 분산 수용되는 방향으로 자리잡고 있는데, 이를 지원하기 위해 유일한 문서만을 식별하여 저장해야 하는 필요성이 대두되고 있다. 하나의 서버내에 중복적으로 존재하는 문서를 제거하는 문제뿐만이 아니라 분산환경에서 다수의 서버에 동일한 문서가 중복저장되어 있다는 사실을 파악하는 문제도 전자도서관의 원활한 서비스를 위해 매우 중요하다. 이러한 기술을 통하여 전자도서관은 대용량의 정보를 효율적으로 관리하므로서 소요되는 비용 및 저장공간을 절약할 수 있다. 따라서 도서관에 존재하는 기존의 문서와 새롭게 추가되는 문서의 중복성을 검사하는 기술의 개발은 절실하다고 할 수 있다.

4. 결론 및 기대효과

인터넷과 멀티미디어 관련 시장은 지속적인 성장이 예측되고 있다. 그러나, 이러한 성장에도 불구하고 영화, 음반, 이미지를 제작/판매하는 멀티미디어 산업은 성행하는 불법복제 때문에 큰 피해를 입고 있다. 따라서 저작권보호기술은 디지털저작물이 폭발적으로 증가하는 추세에 비추어 그 적용분야 및 기대효과는 지속적으

로 증가할 수 밖에 없다고 할 수 있다.

특히, 멀티미디어 디지털 데이터는 인간 생활과 밀접한 관계를 형성하고 있으며, 그 편리성 때문에 인간의 지적 창작물은 점점 디지털 형태로 표현되어 제작되고 있다. 그러나, 디지털 데이터는 쉽고 완전하게 복제될 수 있다는 치명적인 약점을 갖고 있다. 이런 약점 때문에, 멀티미디어 디지털 데이터가 원래의 창작물인지 아니면 복제물인지를 판별해야 하는 경우가 자주 발생하게 되었으며, 보다 근본적으로는 복제물의 무분별한 생산을 억제하거나 설사 복제가 가능하다고 할 지라도 이에 대한 유통을 근본적으로 제한 할 수 있는 방법의 개발이 절실히 필요하다. 이러한 저작권보호기술은 특히 상업적인 분야에서 다양하게 활용될 수 있을 것이다. 즉 콘텐츠의 저작권이 보호되고 불법복제가 억제됨으로써 멀티미디어 콘텐츠의 유통이 정상화되는데 크게 기여할 수 있으며, DOI와 INDECS 시스템과의 연계를 통하여 지적소유권이 인정되는 저작물의 거래내역을 확인하므로서 거래내역의 확인을 통한 거래의 투명성 확보는 저작물의 디지털화를 촉진하고 디지털저작물의 전자상거래를 활성화할 수 있다. 또한, 폭발적으로 증가하는 디지털저작물에 대한 저작권을 기술적으로 보호하는 시스템을 구축하므로서 디지털저작물의 체계적인 관리 및 유통의 효율화를 도모하고 등록된 디지털저작물에 대한 저작권을 보호하므로서 디지털저작물에 대한 저작을 활성화할 수 있다.

결론적으로 현재 정부에서 추진하고자 하는 DOI 시스템(The Digital Object Identifier System)의 요소기술로서 요구되는 분야가 문서 복제판별기술(copy detection)과 워터마킹(watermarking)과 같은 저작권보호기술 분야이다. 미국을 비롯한 선진국에서는 이미 1990년대 중반부터 디지털 저작물에 대한 효율적인 접근과 저작권 보호에 대한 필요성을 인식하고 이와 관련된 연구 및 각종 사업을 적극적으로 추진하고 있다. 따라서 이러한 분야에 대한 보다 적극적인 연구와 과감한 투자가 필요할 것이다.

참 고 문 헌

G. Voyatzis, N. Nikolaidis and I. Pitas, "Digital Watermarking: An Overview," Proc, of EUSIPCO '98, Sep., 8-11, 1998.

H. Garcia-Molina, L. Gravano, N. Shivakumar, "dSCAM: Finding Document Copies across Multiple Databases", Proceedings of 4th International Conference on Parallel and Distributed Information Systems (PDIS'96), 1996.

H. Garcia-Molina, S. Ketchpel, N. Shivakumar, "Safeguarding and Charging for Information on the Internet," ICDE'98. 1998.

J. J. K. O' Ruanaidh, W. J. Dowling, F. M. Boland, "Watermarking digital images for copyright protection," IEEE Proceedings on Vision, Image and Signal Processing, 143(4), pp. 250-256, Aug., 1996.

I. J. Cox, M. L. Miller, "A Review of Watermarking and the Importance of Perceptual Modeling," Proc. SPIE Conf. on Human Vision Electronic Imaging II, Vol. 3-16, pp. 92-99, Feb. 1997.

N. Shivakumar, H. Garcia-Molina "Building a Scalable and Accurate Copy Detection Mechanism," Proceedings of 1st ACM International Conference on Digital Libraries (DL '96) March 1996, 1996.

N. Memon and Ping Wah Wong, "Protecting digital Media content," Communications of the ACM, Vol. 41, No. 7 pp35-43, July 1998

R. B. Wolfgang, E. J. Delp, "A watermarking technique for digital imagery: further studies," Video and Imaging Processing Laboratory, Proceeding of the International Conf. on Imaging Science, pp. 279-287, 1997.

W. Bender, D.Gruhl, N Morimoto, A Lu, "Techniques for data hiding," IBM Systems Journal, Vol. 35, No. 3&4, 1996.

배익성, 김강석, 차의영, "디지털 영상의 저작권 보호를 위한 워터마킹에 관한 연구," 한국정보과학회, 1998, 10.

원치선, "디지탈 영상의 저작권 보호," 정보과학회지 제15권 제 12호, pp. 22-27, 1997. 12. 19