

DOI에 기반한 텍스트 저작권 보호 모델

최순희[†], 이경현[‡]

† 부경대학교 전산정보학과

‡ 부경대학교 컴퓨터멀티미디어 공학부

A Text Copyright Protection Model based on the DOI

Soon Hee Choi[†], Kyung Hyune Rhee[‡]

† Department of Computer & Information Science, Pukyong Nat'l University

‡ Division of Computer and Multimedia Eng., PuKyong Nat'l University

요약

인터넷상에서 출판사들의 디지털 저작물 전자상거래를 위하여 개발된 DOI는 디지털 정보의 유일한 식별자로서, 저작권 처리와 웹상에서 정보를 자동으로 신뢰성 있게 찾아주는 기술이다. 본 논문에서는 DOI에 기반한 텍스트 문서의 저작권 보호 방안을 제안한다. 제안된 기법은 인터넷상에서 불법 복사와 같은 행위로부터 전자출판물을 안전하게 이용할 수 있는 방안이다.

1. 서론

인터넷의 발전이 도서관 및 출판시스템에 전자매체를 사용하도록 하는 변화를 가져왔다. 그러나 전자매체의 특징을 보면 불법 복제물이 원저작물을 대체하는 지적 소유권 침해의 저작권 문제를 갖고 있으므로 개개의 저작자 혹은 출판사들이 디지털 저작물 공급을 인터넷과 같은 공유 네트워크에 공개하기를 꺼리게 하는 현상을 유발시켰다. 따라서 불법 복사와 같은 행위로부터 저작자의 권리를 보호하고 건전한 정보사회를 창출하기 위해 정보보호의 중요성이 크게 대두되었다.

한편, 정보 서비스업자는 정보의 원활한 보급과 이용을 위해서 정보를 자동으로 수집하고, 저작권 문제를 자동으로 처리할 수 있는 구조를 원하게 되었고 이러한 정보 자동 수집과 저작권 자동처리 구축을 위해서는 세계적 표준 기술이 사용되어야 한다. 현재 문서 추적을 위한 기술로써 제시된 기술은 IETF가 제안한 URN(Universal Resource Names)기술이 있고,[1] URN의 제안을 만족하는 기능으로 IDF(International DOI Foundation)의 DOI(Digital Object Identifier)가 있다.[2]

전자상거래를 위한 기술로서 저작권 처리는 저작물

에 권리운영정보를 삽입하여 배포함으로써, 저작물의 저작자와 소유관련 정보를 자동으로 추출할 수 있으며 허가된 사람 이외에는 관련 삽입 정보를 볼 수 없게 워터마킹 처리를 하는 것이다.[3]

따라서 본 논문에서는 출판사들의 디지털 저작물 전자상거래를 위하여 개발된 DOI가 디지털 정보의 유일한 식별자이며, 저작권 처리와 웹 상에서 자동으로 신뢰성 있게 정보를 찾아주는 기술이므로, DOI에 기반한 저작권 보호 방안을 제안한다. 제안된 기법은 텍스트 문서를 대상으로 불법 복사와 같은 행위로부터 인터넷상에서 안전하게 전자출판물을 배포하고 이용할 수 있는 방안이다.

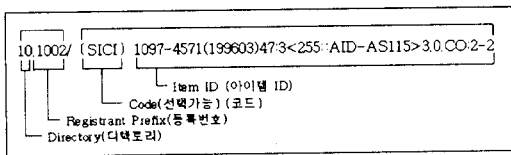
2. DOI

2.1 DOI의 개념

웹상에서 제공되는 디지털 원문에 대한 위치정보는 대개 URL을 이용하지만, 시스템 주소 변경 혹은 대상물의 위치가 변경되면 더 이상 존재하지 않은 URL을 가리키게 되므로 계속적으로 관리하지 않으면 이전 URL을 사용하는 사람들은 원하는 대상물을 얻을 수 없게 된다. 그러므로 URL의 경우는 디지털 대상물에 대한 영구적인 식별자로서의 기능을 제대로 수

행하지 못하기 때문에 W3C의 IETF는 범세계적으로 유일하게 식별가능하고, 영구적인 식별자로서 URN 사양을 정의하였으며, 이 사양에 따른 대표적인 구현 사례로 IDF(International DOI Foundation)의 DOI가 있다. DOI의 역사는 1994년 미국 출판협회가 책에 부여되는 ISBN과 같은 번호 체계를 온라인 자료에 적용하여 인터넷 상거래를 자동화하고, 저작자를 보호하기 위한 시스템 개발을 위해 제안을 하였으며, 그 후 DOI는 1998년 미국 표준화 기구(NISO)가 제안한 안 [4]에 의해 표준화가 된 인터넷 문서의 전자상거래를 위한 디지털 정보의 유일한 식별자이며, 저작권 처리와 웹 상에서 자동으로 신뢰성 있게 정보를 찾아주는 기술로서 인식되었다. DOI는 자원탐색, 상거래, 그리고 저작권관리 세 가지 측면에서 응용이 가능하다.

2.2 DOI 식별자 구조 및 동작법



<그림 1> DOI식별자 구조 사용 예

DOI 식별자 구조는 <그림1>에 주어졌고 이때 Prefix는 해당 DOI 레코드를 관리하고 있는 '디렉토리' 번호와 디렉토리 관리자가 출판물의 저작권을 가진 각 출판사에 부여하는 '등록번호'로 구성되며 즉 10.1002에 해당되며, Suffix는 '/' 이하로 대상물 식별 기호 부분으로, 대개의 경우 BICI, ISWC, ISBN, ISSN 등과 같은 국제 표준 중에서 하나를 사용할 수 있다.[5]

DOI의 동작법 및 등록 에이전시 추진 체계에 관련된 세부적인 내용은 [6]의 논문에 상술된다.

3. 텍스트 문서의 저작권 보호 모델 분석

마킹 기술을 적용하여 인터넷 문서를 배포할 경우, 정당한 각 문서 수령인은 마크가 삽입된 유일한 각각의 문서를 받게 되는데, 각 마크는 텍스트 상에서 눈에 보이지 않도록 교묘하게 처리된다. 문서에 유일한 마크를 추가하는 것은 많은 효과를 제공하는데 가령 예를 들면, 저작권 보호와 같은 어플리케이션에서, 이미지에 삽입된 마크는 저작물의 정당한 수령인에 대한 "식별자"로서 사용될 수 있으며 또한 마크가 삽입된 문서를 분석했을 때, 그 마크의 저작물이 정당한 수령인의 것 일수도 있으며, 불법 복제물이라고 판단 되었을 경우에는, 누구로부터 불법 복제되었다는 것을

"추적" 할 수 있다.

3.1 마킹기술

1) 줄 이동 코딩

이 마킹 기술은 인접 텍스트 줄을 수직으로 옮겨놓은 것에 의해 마크를 삽입한다. 줄은 위 또는 아래로 움직이고, 한편 인접한 한 줄 위 또는 아래 줄은 이동시키지 않는다. 이런 이동되지 않은 인접한 줄들은 부호화 과정에서 참고 위치(reference location)로 취급된다. 워터마크는 지울 수 없고, 그리고 자료에 워터마크가 표시되도록 텍스트의 일부분에 역시 추가할 수 있고 판독에 있어서 변하지 않으므로, 이 기술은

2) 단어이동코딩

이 코딩 기술은 텍스트 라인 안에 있는 단어의 위치를 수평이동 시켜 삽입한다. 단어는 왼쪽 또는 오른쪽으로 이동되고, 직접적으로 인접한 단어는 이동시키지 않는 반면, 인접한 두 단어의 가운데 단어는 왼쪽 또는 오른쪽으로 이동된다. 이동되지 않는 단어들을 참고위치로 취급할 수 있다.

3) 문자코딩

문자코딩은 개개의 문자의 특별한 특징을 변경하여 마크를 삽입하는 기술의 종류로서, 변경 가능한 특징의 예들에서 개개의 문자들의 높이, 또는 다른 문자들과 관계 있는 위치 변경을 포함한다(예들 들면 "kerning"(자간 좁힘)을 조정). 문자 변경에 의해 마크를 눈에 보이지 않도록 삽입하고자 할 때는 변경될 수 있는 문자의 전후관계를 주의 깊게 처리해야 한다. 마크의 검출시에는 마킹 기법과 오리지널 문서로부터 정보가 필요하거나 필요하지 않을 수도 있다.[7]

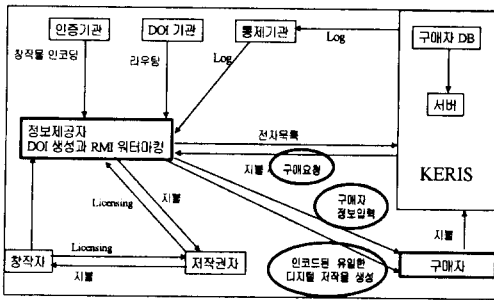
4. DOI에 기반한 텍스트 저작권 보호 모델

제안 및 구현

4.1 DOI에 기반한 텍스트 저작권 보호 모델

제안

본 논문에서 제안하는 저작권보호 모델은 KERIS의 상거래 모델 부분에서 정보공급자에 의한 DOI 정보와 구매자 정보를 함께 워터마크 정보로 삽입하여 유일한 디지털 저작물을 생성시킨 후 각각의 구매자에게 배포하는 방식으로 <그림 2>에 나타내었다.



<그림 2> KERIS 상거래 모델에서의 DOI 기반 저작권 보호 모델 제안

<그림 2>에서는 KERIS의 상거래 모델 설명과 DOI에 기반한 저작권 보호 모델 제안 설명을 동시에 하고 있다. 먼저 KERIS 상거래 모델에 대하여 설명하면, 저작자는 그의 저작물에 RMI(Right Management Information)를 넣는다. 저작물은 저작자/저작권 소유자, IP(Information Provider) 사이에 면허 허가를 통하여 작품을 IP에게 제공한다. DOI 이름 부여 기관인 IP는 각 작품에 DOI 번호를 부여하며, 구매자가 원할 때 RMI에 면허 정보를 추가하고 워터마킹 한다. 작품의 전자목록이 IP에 의하여 생성된다.

IP는 CA(인증기관)에 의해 제공되는 개인키를 가지고 공개키 암호화 알고리즘을 사용하여 전자목록을 암호화한 뒤, ISP(여기서는 KERIS)로 올린다. ISP는 전자목록을 모으며, 회원(구매자) DB를 가지고 있다. 만약 구매자가 특정한 작품을 필요로 하면, ISP는 DOI번호를 DOI 기관으로 보내고, DOI 기관은 해당되는 IP에게 요구를 보낸다. IP 해당 기관은 해당 작품을 구매자에게 제공한다. 이러한 거래 과정은 통계 기관에 저장되며, IP와 저작권 소유자에게 보고된다. 이러한 구조는 정확한 로얄티 분배를 증명하는 역할을 제공한다.[8]

다음은 본 논문에서 저작권 보호를 위하여 모델을 제안하는 부분으로 <그림 2>에서 진하게 부각시켜 설명하고 있다. 디지털 저작물 구매 희망자와 해당 정보공급자가 연결되면, ① 구매 희망자가 구매를 요청하고, ② 정보제공자는 구매자에게 구매자 정보 및 신용카드 번호를 물어서, ③ DOI 정보에 신용카드 정보를 추가하여 함께 워터마크 된 유일한 디지털 저작물을 만들어 그 구매자에게 전송한다.

여기서 DOI 식별자와 구매자 정보는 KERIS의 DOI 표기법을 따랐으며 <그림 3>에 나타내었다.

```

KERIS_Number := <DIR>.<REG>/<DSS>
<DSS> := <Name>/<RightHolder>
<Name> := [<ISDI>] | <Serial>
<ISDI> : (ISBN | ISSN | SICI ) <Number>
<Serial> := <N><Y>-<D1>.<D2>
    <그림 3> KERIS DOI 표기법
    
```

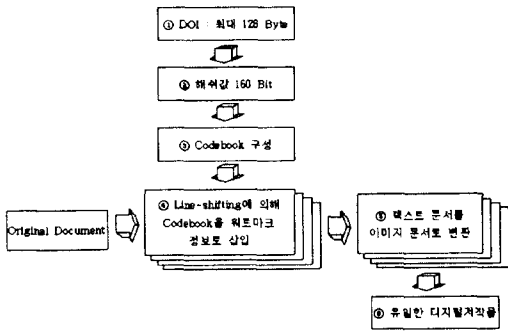
먼저 DOI 식별자 중 디렉토리는 10으로 정의된 DOI Server로 한다. 창작자와 배포자 식별자는 ISBN의 번호 체계를 따라 표현하여 사용하고, 구매자 정보는 신용카드 번호를 사용하고자 한다. 신용카드 번호는 총 16자리로써 세계적으로 중복되지 않게 번호가 부여되고 있을 뿐만 아니라 소유자가 다른 사람에게 자신의 신용카드 번호를 일반적으로 알려줄 것이라고 판단되지 않으므로 선택할 수 있다. 디지털 저작물에 대한 DOI 식별자 구성을 <그림 4>에 나타내었다.

DOI = 10.1064/[ISBN]89-7155-033-6/8309 2309 9253 0001

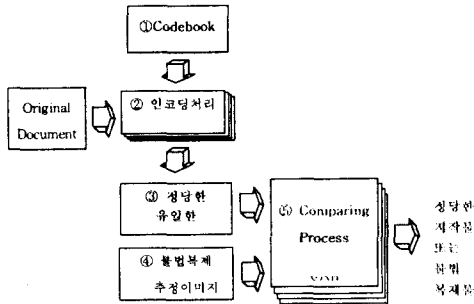
<그림 4> DOI 구성

<그림 2>의 ③과정을 <그림 5>에서 보다 상세하게 나타내었고, 이를 단계별로 살펴보면 다음과 같다.

- ① DOI 정보로써 최대 128Byte까지 넣을 수 있으나 본 논문에서는 총 47Byte인 디렉토리 서버번호, 배포자 번호, 배포자료에 대한 식별자, 구매자 정보가 연결되어 들어가며 나머지는 0으로 padding 된다.
- ② 128Byte DOI 정보를 해쉬하여 160 bit의 이진 값으로 변환한다.
- ③ ②의 결과는 Codebook을 구성한다.
- ④ 원본 문서에 줄 이동 마킹 기법을 이용해서 ②의 DOI에 대한 해쉬 값을 워터마크 정보로 삽입한다.
- ⑤ 텍스트 문서를 이미지로 변환함으로써, 유일한 디지털 저작물(⑥)을 만든다.



<그림 5> 제안 모델의 워터마크 과정



<그림 6> 제안 모델의 워터마크의 검출 과정

<그림 6>은 워터마크의 검출과정으로서, 불법복제물로 추정되는 디지털 저작물이 발견되었을 때 정당한 디지털 저작물인지를 원본과 비교하여 검출하는 과정이다.

- ① 우선 의심되는(Suspected Image) 전자출판물의 진위여부를 판단하기 위해 Codebook을 이용한다.
- ② 원문서는 Codebook의 정보로 Encoding처리한다.
- ③ ②로써 믿을 수 있는 전자출판물을 생성한다.
- ④ ⑤의 의심되는 출판물의 Gif file XOR 비트 연산으로 비교하여, 일치여부로서 정당한 저작물인가를 판정한다.

4.2 제안 모델의 구현 예

앞 절에서 제안한 워터마크 삽입과정과 검출과정을 위한 시스템 환경은 IBM-PC Pentium(166MHz, RAM 32MB)과 Microsoft Windows 95 운영체제를 사용하였으며, 프로그래밍 언어는 자바 SDK 1.2.2를 사용하여 구현하였다.

1) 워터마크 삽입 과정 및 결과

<단계 1>, 구현에서 사용할 텍스트 저작물은 인터넷에서 영문자료 한 편중 일부를 캡처 하였다.

<단계 2>, 텍스트 문서와 문서에 삽입시키고자 하는 정보인 DOI를 입력 값으로 하여 인코딩 처리를 한다. DOI 정보는 제안 모델에서의 사용 예 <그림 4>를 사용할 것이다.

<단계 3>, DOI 정보에 대한 해쉬 함수는 SHA-1 알고리즘을 사용하여 160bit 값을 구한 후 Codebook을 생성한다.

<단계 4>, 인코딩 처리는 줄 이동을 사용하고 있으며 각기 구현하는 하드웨어 환경에 따라 다를 수는 있으나, 해상도 640×480에서 1/640의 단위(1 pixel)로써 대상 줄을 이동하였다. 이동대상 줄 선정에 관한 처리는 DOI 정보로써 구한 160bit의 hash값에 따라 결정되었다.

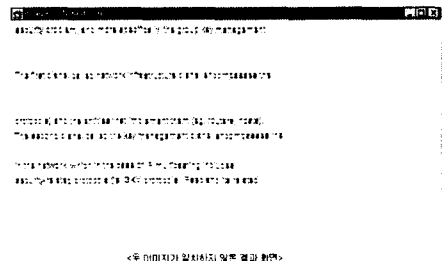
<단계 5>, 줄 이동 처리가 된 텍스트 문서는 한 페이지단위로 640×480 크기의 GIF 이미지 파일로 변환되어 저장된다. 이 파일이 유일한 디지털 저작물로 출력된다.

2) 워터마크 검출 과정 및 결과

워터마크 검출과정은 불법 복제물로 추정되는 디지털 저작물이 발견되었을 때, 정당한 저작물인지를 원본과 비트 단위 Exclusive-OR(XOR)연산으로써 진위 여부를 판정할 수 있도록 구현하였다.

- ① 진위여부가 가려지지 않은 문서를 '불법복제 추정 이미지'라고 한다.
- ② 그에 해당하는 출판사의 Database에 저장되어 있는 해당 Codebook 파일에서 인코딩 되기 전의 오리지널 텍스트 문서에 인코딩 처리를 한다.
- ③ 정당한 이미지, 즉 Unique Image를 생성한다.
- ④ 정당한 이미지와 불법복제 추정 이미지를 XOR 연산하여 그에 대한 결과를 화면상에 출력하여 그 진위여부를 판정한다.

예를들어, 불법복제 추정 이미지와 정당한 유일한 이미지를 XOR 연산 후 결과 화면이 공백으로 나타나면 정당한 이미지로 판단할 수 있다.



<그림 7> 두 이미지가 일치하지 않은 결과

불법복제 추정 이미지와 정당한 유일한 이미지를 XOR 연산 후 결과 화면이 <그림 7>과 같다면 불법적인 복제물로 판단할 수 있다. <그림 7>에서 출력된 줄들은 줄 이동하기 위한 줄 선택이 서로 다르기 때문에 XOR 연산 후 사라지지 않고 비정상적으로 출력된 형태이다.

5. 결론

컴퓨터, 디지털기술, 정보통신기술의 발전으로 도서관 및 출판시스템 전자매체의 사용이 진전되어져 오고 있다. 출판사들은 인터넷과 같은 공유 네트워크에 디지털 저작물을 배포하는 데 있어 여러 가지 문제 중 최근 부각되고 있는 불법 복제로부터 저작권을 보호할 수 있는 구조 구축을 희망하고 있으며 이미 구현을 위해 DOI를 활용할 수 있다. DOI는 전자상거래를 위하여 ISBN과 같은 저작물의 정보를 식별하고 자동 추적할 수 있는 표준 정보 식별자로써 저작물의 창작자 정보, 출판사 정보에 구매자 정보를 추가하여 정당한 문서 수령인 외에는 볼 수 없게 워터마킹 처리를 할 수 있는 구조이므로 모든 형태의 저작물에 적용할 수 있고 불법 복제로부터 추적도 할 수 있어 저작권 보호에 효율적으로 활용할 수 있다.

본 논문에서는 저작물의 저작자와 소유관련 정보를 삽입하는 워터마크 기술과 검출과정을 분석하여 DOI를 기반으로 저작권을 보호할 수 있는 모델을 제안하고 구현하였다. 제안된 모델은 정보공급자에 의한 DOI 정보와 구매자 정보를 해쉬값으로 160bit의 이진값으로 변환시켜 Codebook을 구성하여 오리지널 문서에 줄 이동을 이용해서 DOI에 대한 해쉬값을 워터마크 정보로 삽입하여 유일한 디지털 저작물을 생성시키도록 했다.

제안된 모델은 전자출판으로의 변화를 위한 자생력 있는 기술로서 성립될 수 있으며, DOI가 할당화 되면 주로 전자출판물, 과학기술 의료 및 학술·논문지 위주로 시장이 조성될 것으로 예상되고 있으며, 이미 국내에서도 DOI를 자체 도입하여 활용할 계획을 갖고 있는 정보회사도 존재하고 있다. DOI의 효용성은 저작권 정보 제공 및 보호뿐만 아니라 위치 정보와 홍보 효과 측면에서 일반 ISBN과는 큰 차이가 있기 때문에 DOI 체제가 본 계도에 들어서면 인터넷상에서의 체계적인 정보검색과 접근의 용이함으로 인하여 사용자에게 희망하는 정보를 보다 폭넓게 제공할 수 있게 될 것이고, 디지털 도서관의 정보서비스 및 구축에도 크게 기대할 수 있는 것으로 판단된다. 추후 연구로써 앞서 설명된 다양한 마킹 기술을 조합함으로

써 삽입된 워터마크 정보로부터 유일의 DOI 정보를 추출해 낼 수 있는 기법에 대한 연구를 수행하고자 한다.

[참 고 문 헌]

- [1] K. Sollins, Functional Requirements for Uniform Resource Names, <ftp://ftp.isi.edu/in-notes/rfc1737.txt>.
- [2] Norman Paskin, "DOI : Current Status and Outlook", D-Lib Magazine, May 1999.
- [3] Peter Lambert Tagish, Imprimatur ByLine Trial Report, IMP/5060 B, Jan 1999.
- [4] NISO, Syntax for the Digital Object Identifier, ANSI/NISO Z39.84-XXX, 1998.
- [5] 한혜영, "국내 학술지 게재논문을 위한 SICI기반 DOI 체계, 이화여대문헌정보학과 기념논문집, 1999.
- [6] 이창열, "DOI 관련 기술 동향과 국가 에이전시의 역할", http://www.kpa21.or.kr/w_doi.htm.
- [7] J. T. Barssil, S. Low, N. F. Maxemchuk, "Copyright Protection for the Electronic Distribution of Text Documents" Proceedings of the IEEE, June 1999.
- [8] 이창열, "전자상거래 기반 구축과 비즈니스 모델 개발", http://www.kpa21.or.kr/w_doi.htm.