

# NewsML과 UCI를 적용한 뉴스 콘텐츠의 온라인 유통모델

박창신\*, 김 덕\*\*

\*조선일보 경영기획실, \*\*조선일보 경영기획실

## A New Online News Service Model, based on NewsML and UCI Systems

Park, Chang-Shin Kil, Duke

Chosun Ilbo, Chosun Ilbo

E-mail : heri@chosun.com

### 요 약

News contents, produced for paper readers, are more and more being used online instead of offline. Internet sites, especially portals(naver, daum, nate etc.) are dominant marketplaces, where news are exchanged and values are added. So, establishing a new online news service system, which can satisfy news provider(copyright owner) and internet service provider together, is a necessary task under current online-dominant news service environment.

UCI(Universal & Ubiquitous Content Identifier) and IPTC NewsML(News Mark-up Language) are considered as useful standards to compromise protection of news-copyright and enhancement of online use of news contents. This study is based on a real case of 'NewsBank' in Korea. We expect that this study can show an inspiration to obtain two contradictory goals of copyright protection and free online use of copyright.

### 1. 서론

본래 뉴스(News)는 누군가 읽거나 봐야 효용을 발휘한다. 아무도 읽지 않고 보지 않는 뉴스는 그저 개인의 비망록(備忘錄)에 불과할 것이다.

오늘날 디지털화된 뉴스 콘텐츠는 온라인에서 상당부분 소비된다. 따라서 뉴스의 온라인 유통은 오프라인 유통에 못지않게 중요해졌지만, 이에 대한 인식과 사회적 인프라는 전반적으로 미흡한 게 현실이다. 현재 다수의 언론사들은 네이버·다음·네이트 등의 포털사이트에 뉴스 콘텐츠를 공급하지만, 분류체계와 전송방식 등이 제각각이다.

뉴스 콘텐츠는 시대와 수용자의 요구에 의해 보다 다양해지고 있으며, 방대한 규모의 과거와 현재

의 뉴스 콘텐츠를 효율적으로 관리하고 이용을 활성화하기 위해서는 우선적으로 분류체계와 식별부호를 포함한 전송방식의 통일이 선행되어야 할 과제가 아닐 수 있다. 콘텐츠 생산 단계부터 정교하고 표준화된 분류 및 식별체계가 적용되고, 이런 체계에 의해서 뉴스 콘텐츠가 유통될 때 소비자는 보다 쉽고 편하게 원하는 뉴스 콘텐츠를 검색 등을 활용해 이용할 수 있는 등 이용활성화는 물론 저작권의 기술적 보호조치 등이 가능하기 때문이다. 이에 국가 표준 디지털 콘텐츠 식별체계인 UCI(Universal & Ubiquitous Content Identifier)<sup>1)</sup>

1) UCI는 인터넷상에서 디지털 콘텐츠의 효율적인 유통과 활용을 위해 디지털 콘텐츠 하나하나에 유일한 코드를 부여하고 이를 관리해 주는 식별체계다. UCI는 구문구조, 메타데이터, 운영절차, 운영시스템이 결합된 체계이며, 기존의 KOI(Knowledge Object

를 뉴스 콘텐츠의 온라인 유통에 적용함에 있어, 분류 및 전송체계를 어떻게 정립해야할 지를 고민해야 한다.

현재 AFP, AP, 로이터, 교토통신 같은 세계 유수의 뉴스 통신사는 물론 BBC, 뉴욕타임스 일본신문협회 등 전 세계 다수의 언론사들은 IPTC(국제언론통신평의회, International Press Telecommunications Council)<sup>2)</sup>를 구성하고, IPTC가 정한 NewsML<sup>3)</sup>의 분류체계를 디지털 콘텐츠 거래의 표준으로 채택하고 있다. 이렇게 디지털 뉴스 콘텐츠의 세계적 유통에 있어 NewsML 분류체계가 사실상의 세계 표준으로 널리 적용되고 있는 가운데, 우리나라에서도 한국언론재단이 지난 7월 26일 IPTC 분류체계를 한글화한 '한국형 NewsML 표준안(KNF)'을 내놓은 바 있다. 또한 2007년 조선일보, 동아일보, 한국일보, 한국경제 등 10여개 신문사들은 '뉴스뱅크'<sup>4)</sup>라는 협의회를 구성하여 IPTC NewsML 분류체계를 적용한 공동의 뉴스콘텐츠 아카이브를 구축했다. 뉴스 콘텐츠에 UCI를 적용하려면 IPTC의 틀을 고려할 수밖에 없는 것이 현실이다.

본 연구에서는 UCI와 IPTC NewsML 체계를

Identifier, 과학기술정보 식별체계), DOI(Digital Object Identifier, 미국출판협회 식별체계), ISBN(International Standard Book Number, 국제표준도서번호) 등 다양한 온·오프라인 식별체계를 수용할 수 있다. 정보통신부 산하 한국정보사회진흥원의 주도하에 개발됐으며, 2004년 특허 출원에 이어 2006년 6월 국내 TTA(Telecommunications Technology Association, 한국정보통신기술협회)의 표준화 절차를 밟아 현재 이를 국가정보통신표준(KICS)으로 채택하기 위한 절차를 진행 중이다.

- 2) IPTC는 1965년 AEAP(the Alliance Europeenne des agences de Presse), ANPA(현재의 NAA), FIEJ(현재의 WAN), 아메리카뉴스연합(미국과 캐나다의 조인트 협의회) 등이 설립하였으며, 1970년대부터 본격적으로 뉴스의 국제 표준화 작업을 추진하였다.
- 3) NewsML(News Mark-up Language)은 XML로 쓰여진, IPTC가 발표한 국제 표준 뉴스 포맷이다.
- 4) '뉴스뱅크'는 2007년 초 결성된 언론사들의 사업 연합체이다. 2007년 10월 현재 국민일보, 동아일보, 문화일보, 세계일보, 스포츠조선, 전자신문, 조선일보, 한국경제, 한국일보, 헤럴드미디어 등 10개 신문사가 참여하고 있다. 당초 뉴스뱅크는 조선일보가 NewsML을 적용한 뉴스 아카이브 시스템을 구축하고, 이를 각 언론사에 제공함으로써 '언론사 공동의 뉴스 아카이브 사업'의 형태로 성립되었다.

온라인의 뉴스콘텐츠 유통체계에 조화롭게 적용하는 방안을 모색하고, 이를 위한 주요 언론사들의 노력과 성과를 소개하고자 한다.

## 2. 본론

### 2.1. 한국판 NewsML체계의 정립

IPTC NewsML 분류체계의 국내 도입을 위해서는 NewsML의 한글화 뿐 아니라, 분류 체계를 한국의 상황에 맞게 수정 및 확장이 필요했다. 이에 조선일보, 동아일보 등 뉴스뱅크<sup>5)</sup> 소속의 10개 언론사들은 2006년 9월부터 2007년 8월 말까지 NewsML의 한글화 및 수정·확장 작업을 진행하였다.

IPTC는 시스템 관리 혹은 시스템간 인터페이스를 위해 통제되는 매체(MEDIA), 형식(FORMAT) 등과 같은 토픽셋(TopicSet) 외에도, NewsML의 국제적 통용을 위해서 NewsML의 일부 토픽셋에 대해서는 뉴스제공사가 독자적으로 관리할 수 있도록 했다. 정당(Party), 인물(People), 회사(Company), 발생지(Location), 기관(Organization), 상품(Product) 등이 이에 속하며, 필요할 경우 뉴스 제공사가 새로운 Topicset을 만들 수도 있다.

토픽셋의 하나인 주제분류(SubjectCode)의 경우, 17개의 대분류에 더해 18번째로 '북한'을 첨가하고, 1400개의 세부 코드에 더해 62개의 코드를 추가했다. 뉴스 제공사(provider)의 경우 31개의 코드를 더했고, 사진의 경우 미국사진기자협회가 정한 7개의 장르를 추가하는 등 보완 작업을 했다. 또한 한국의 상황에 맞춰 IPTC에는 없는 15개의 새로운 토픽셋을 추가했다. 이렇게 해서 완성된 토픽셋은 다음과 같다.

- 5) '뉴스뱅크'는 2007년 초 결성된 언론사들의 사업 연합체이다. 2007년 10월 현재 국민일보, 동아일보, 문화일보, 세계일보, 스포츠조선, 전자신문, 조선일보, 한국경제, 한국일보, 헤럴드미디어 등 10개 신문사가 참여하고 있다. 당초 뉴스뱅크는 조선일보가 NewsML을 적용한 뉴스 아카이브 시스템을 구축하고, 이를 각 언론사에 제공함으로써 '언론사 공동의 뉴스 아카이브 사업'의 형태로 성립되었다.

◇ 신뢰도	◇ 긴급도	◇ 전달 우선 순위
◇ 중요성	◇ 메타데이터의 중요성	◇ 독자와의 관련성
◇ 공개 지시	◇ 주제 코드	◇ 포맷
◇ 주식	◇ 뉴스 라인 타입	◇ 마일(MIME) 타입
◇ 미디어 타입	◇ 토픽 타입	◇ 메타데이터
◇ 뉴스 아이템 타입	◇ 롤(ROLE)	◇ Property
◇ 전달 방법	◇ 독자층	◇ 장르
◇ Newsletter 의 라벨	◇ Product	◇ 언어
◇ 국가명	◇ 통화(단위)	◇ 단체명*
◇ 지역명	◇ (M)산업분류코드	◇ 나스닥 회사 코드

<표 1> IPTC가 정한 토픽셋

①섹션(section): 13, ②면(pagetype):101, ③인물(people): 4834, ④지역(region):5605, ⑤기관(organization): 113, ⑥회사(company): 2642, ⑦상품(product):488, ⑧감각정보(senseinfo): 238, ⑨동작정보(actioninfo):279, ⑩물체정보(objectinfo):73, ⑪배경정보(backgroundinfo):128, ⑫신상정보(bodyinfo): 175, ⑬직업(jobcode):188, ⑭그래픽정보(graphicinfo):6, ⑮사진장르(photogenre):8

<표 2> 뉴스뱅크가 추가한 토픽셋\*

가장 활용도가 높은 토픽셋인 주제 코드(SubjectCode)는 8자리(○○|○○○|○○○ 대·중·소분류)의 숫자로 이루어진 1400여개의 세부적인 '주제 코드'를 포함하고 있다. 현실적으로 뉴스콘텐츠의 온라인 유통에 있어서는 각 콘텐츠에 따라 붙는 메타데이터 가운데 주제분류가 가장 중요하고 우선적으로 활용된다. 따라서 뉴스콘텐츠의 온라인 유통에 UCI를 적용하려면, 먼저 이 주제코드 체계부터 정립되어, 뉴스콘텐츠 생산자와 유통사업자가 이를 실제 서비스에 적용해야 한다.

6) 토픽셋 옆에 써놓은 숫자는 각 토픽셋별로 세부 항목의 숫자를 뜻한다. 예를 들어 신문의 지면정보(정치면, 사회면, 경제면 등)는 101개로 구분해놓았다. 인물은 신문에 자주 등장하는 유명인사 4834명으로 인물 토픽셋을 형성했다. 물론, 이들 토픽셋과 세부항목은 언론사가 얼마든지 필요에 따라 가감해서 사용할 수 있다.

Name	Description_ko	Name	Description_ko	Name	Description_ko
01000000	문화	01011001	클래식음악	01018001	무형문화재
01001000	고고학	01011002	기악	01018001	NA
01002000	건축	01011002	포크음악	01018002	유형문화재
01003000	투우	01011003	성악	01018003	청원기념물
01004000	축제/이벤트	01011003	저즈음악	01019000	관습/전통
01004001	문학제	01011004	악기	01019004	관혼상제/예법
01005000	영화	01011004	대중음악	01019005	민간관습
01005001	공상과학영화(SF)	01011005	음악원형/음악	01020000	예술(일반)
01005001	영화제	01011005	컨트리음악	01020001	민속예술
01006000	무용	01011006	록큰롤 음악	01021000	엔터테인먼트(일반)
01006001	발레	01012000	회화	01021001	어워드(엔터테인먼트분야)
01007000	패션	01012001	공예	01022000	문학(일반)
01007001	복식(액세서리)	01012002	미술공모	01022001	문학교류(한류)
01008000	언어	01012003	미술전시/전람회	01022001	문학발견/문학면담
01008001	영어	01012004	서예	01022002	문학시설
01008002	외국어	01013000	사진	01023000	나이트클럽
01008003	한국어(한글)	01014000	라디오	01024000	만화
01009000	도서관/박물관	01015000	조각/소조	01025000	애니메이션

<표 3> 뉴스콘텐츠의 주제코드 분류표\*

뉴스콘텐츠의 표준 주제코드체계는 합리적인 온라인 콘텐츠 유통의 출발이다. 이렇게 완성된 코드체계는 다른 토픽셋과 함께, 개별 뉴스 콘텐츠의 메타데이터를 형성하며, 온라인 유통에 있어서 다양한 용도로 활용된다.

## 2.2. NewsML과 UCI의 연동

2007년 10월 현재 뉴스뱅크의 뉴스 아카이브 시스템에는 자체 개발한 CMS(Content Management System)가 가동되고 있으며, 여기에는 앞서 설명한 주제코드분류체계(subjectcode)를 포함하여 뉴스뱅크의 토픽셋 체계가 적용되고 있다. 이 체계는 IPTC NewsML체계를 수용하여 한국 상황에 맞게 보완한 것이어서, 기본적으로 국제 표준이다.

뉴스뱅크시스템이 CMS는 다시 UCI의 RA(Registration Agency, 등록관리기관) 서버가 연결되어 있고, 뉴스뱅크의 RA서버는 한국정보사

7) <표 3>은 IPTC의 토픽셋 중 주제코드분류체계(subjectcode)를 한글화하되, 본 논문에서 밝혔듯이 '북한'을 18번째 대분류로 추가하고, 중/소분류 코드에 62개 코드를 추가하여 완성한 것이다. 전체 주제코드분류체계는 01000000(문화)부터 18012000(재외/동포)에 이르기까지 1400여개 항목으로 구성되어 있으며, 본 논문에서는 극히 간소화한 형태로 앞 부분만을 예시한 것이다.



회진홍원이 관리하는 UCI 총괄시스템인 루트서버 (Root서버)로 연결되어 있다.

IPTC 표준 체계로 짜여져 있는 뉴스 아카이브 시스템에 UCI의 유통표준을 적용하기 위해서는 각 뉴스콘텐츠에 표시되는 메타데이터의 포맷을 이해할 필요가 있다. 이와 관련해 IPTC 표준 체계에서는 모든 콘텐츠에 관리적 메타데이터 (Administrative Metadata), 권한에 관한 메타데이터 (Rights Metadata), 설명적 메타데이터 (Discriptive Metadata)를 필수적으로 명기하도록 했다. 뉴스뱅크 아카이브 시스템도 이 규약을 준수하고 있다.

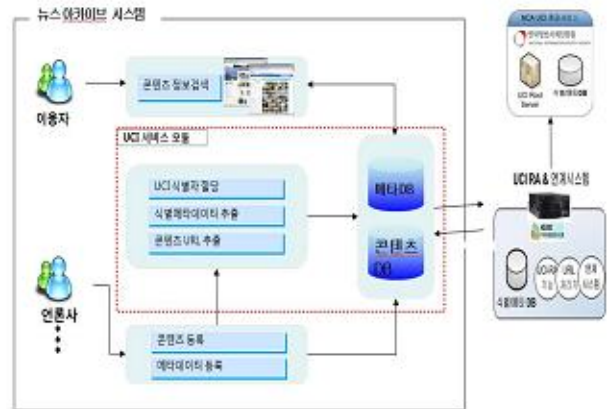
이에 비해 한국정보사회진흥원이 UCI 식별체계에서 필수적인 표준 식별메타데이터 항목으로 정한 것은 identifier(기존 식별자), title(자원명), type(유형), mode(표현형태), format(파일형식), contribution(기여), contributor(기여자), contributorRole(기여자역할) 등이다. UCI의 식별 메타데이터 항목에 NewsML을 적용한 뉴스콘텐츠 메타데이터를 연계시키면 다음과 같다.

UCI 식별메타데이터 (표준)	뉴스콘텐츠 메타데이터	비고
1 UCI코드	자체 일련번호	UCI코드로 변환하거나, 추가로 생성함.
2 identifier(기존식별자)	없음	
3 title(자원명)	AdministrativeMetadata RightsMetadata DiscriptiveMetadata	그 외 필요에 따라 개방 가능
4 type(유형)	Digital, Physical	default값으로 입력됨
5 mode(표현형태)	Visual, Audio, AudioVisual	
6 format(파일형식)	xml, jpg, gif, png, swf	
7 contribution(기여)		8,9의 상위요소
8 contributor(기여자)		
9 contributorRole(기여자역할)	Creator, Writer	

<표 4> UCI 식별 메타데이터에 해당되는 NewsML 콘텐츠의 메타데이터

위 표에서 보는 바와 같이, 뉴스콘텐츠의 주제 분류값은 UCI식별 메타데이터의 항목에 포함되어 있지 않다. 그렇지만, 각 언론사가 각 콘텐츠에 고유한 UCI코드를 부여하여 관리할 경우, 온라인 유통에 있어서 자체 뉴스아카이브의 콘텐츠 메타

데이터에 기록된 분류값을 다양하게 활용함으로써 콘텐츠 유통의 부가가치를 높일 수 있다.



<표 5> 뉴스아카이브의 UCI 연동방안

<표5>는 뉴스 아카이브 시스템에 UCI서비스 모듈을 탑재함으로써, 주제코드를 비롯하여 뉴스 아카이브의 각종 메타데이터를 심분 활용할 수 있도록 설계한 구성도이다.

뉴스 아카이브 시스템에 입력되는 모든 콘텐츠는 콘텐츠 관리 시스템(CMS)에 의해 고유한 번호(CMS Code)가 부여된다. 아카이브 내부의 콘텐츠 관리를 위해 개별 콘텐츠의 Key값을 형성하는 것이다. CMS코드는 뉴스 콘텐츠를 생산하는 각 언론사가 자유롭게 규칙을 정해 부여할 수 있도록 했다.

아카이브 내부에서는 CMS에 의해 콘텐츠가 관리되지만, 일단 해당 콘텐츠가 외부로 유통될 경우에는 반드시 반출시스템(일종의 전송서버)을 거치도록 했다. 반출시스템의 역할은 반출과 함께 각 콘텐츠에 UCI코드를 형성하고, 이를 뉴스아카이브 시스템에 연동되어 있는 UCI-RA시스템에 등록하는 것이다. UCI코드는 <RA코드><CMS 키><전송처 코드>로 구성하였다. 뉴스콘텐츠의 경우 RA코드는 'G010'이다.<sup>8)</sup> 뉴스뱅크시스템에 적용한 CMS키값은 'A0242007060800343480' 식의 20자리로 구성되어 있다. 따라서 UCI코드는 G010

8) UCI코드는 지역을 대표하는 R, 기관에 부여하는 I, 그리고 한 분야를 대표하는 G로 구분된다. 뉴스 콘텐츠에 있어서는 <G010>이 할당되어 있는 상태이다.

에다 하이픈(-)을 더하고 CMS코드 20자리를 합쳐 총 24자리로 구성했다.

외부 반출단계에서 형성한 UCI코드에는 전송처를 추가로 표시하도록 했다. <:XXX>식으로 전송처의 명칭(예를 들어 <:naver>)을 적어 넣을 수 있다. 여기에 덧붙여 전송처의 URL정보도 UCI-RA시스템에 등록하여 관리할 수 있다. 이렇게 하는 이유는 동일한 콘텐츠이더라도 여러 전송처에 동시에 배포될 수 있고, 하나의 전송처이더라도 콘텐츠 제공자와 콘텐츠 사용자(전송처) 사이에 맺어질 계약에 따라 콘텐츠의 이용범위가 특정한 인터넷주소(URL)로 제한될 수 있기 때문이다. 이는 언론사들이 네이버, 다음, 네이버 등의 인터넷 포털기업과 개별적으로 계약을 체결하면서, 콘텐츠 이용범위 등에 있어서 제각각의 조건으로 콘텐츠를 공급하고 있는 점을 반영하기 위한 것이다. 예를 들어 언론사가 특정 포털에 대해 "콘텐츠 사용은 뉴스사이트로 제한한다"는 식으로 계약을 맺는 경우가 있다.

이에 따라 뉴스 아카이브 시스템에서 관리되는 콘텐츠가 전송서버(반출시스템)를 통해 외부로 전송될 경우, UCI-RA서버에 전송처 코드(XXX, YYY 또는 ZZZ)는 물론 전송처의 URL까지 남게 된다. 이렇게 하면, 저작권자가 향후 DRM을 적용하여 특정 콘텐츠가 이용범위 밖에서 사용되었는지 등을 손쉽게 파악할 수 있게 된다.

### 3. 결론

본 연구는 UCI 기반 온라인 유통에 있어서, IPTC 표준 체계로 뉴스 콘텐츠의 주제코드체계를 비롯한 토픽셋을 정립하고, 이 체계가 적용된 뉴스 아카이브 시스템과 UCI시스템을 연동하는 방안을 간단히 제시하였다. 뉴스 유통에 있어서 분류체계가 중요하다는 사실은 굳이 부연이 필요 없을 정도이지만, 다수의 언론사가 각자의 방식으로 뉴스 콘텐츠를 생산해 포털사이트 등의 온라인으로 유통하는 현실에서 분류체계 통일은 결코 쉬운 일이 아니다. 또 표준화된 분류 및 전송체계에 표준화된 식별체계를 적용하는 것은 더욱 어려운 일이다. 그렇지만, 이 일이 어렵다고 해서 회피해서는 뉴스콘

텐츠의 온라인 유통에 있어서 사회적인 비용의 낭비와 온라인 이용활성화의 걸림돌을 제거할 수 없다. 개별 뉴스콘텐츠에 고유한 일련번호를 부여하고, 각 콘텐츠의 몇 가지 메타데이터와 URL 등을 UCI시스템에서 통합 관리하며, 이를 통해 온라인 광고 등의 뚜렷한 사업모델이 형성될 때 온라인 뉴스콘텐츠 유통의 바람직한 모델이 형성될 수 있다. 2007년 하반기 들어 언론사들과 포털사이트간의 뉴스콘텐츠 이용에 따른 갖가지 갈등은 양측이 이런 노력을 통해 양측이 합의할 수 있는 '윈윈(win-win)'의 협력사업을 모색하여 구체화했을 때 해소될 수 있다. 이와 관련해 뉴스뱅크 10개 신문사들이 자발적인 협력을 통해 UCI와 IPTC NewsML의 연계에 의한 뉴스콘텐츠의 새로운 온라인 유통체계를 제시한 점은 의미있는 시도가 아닐 수 없다. 모쪼록 이런 노력이 UCI의 보급과 함께 온라인 유통 시장의 합리적 발전에 기여하기를 기대해 본다.

### [참고문헌]

- [1] 이성숙·송인석(2005), "NewsML을 고려한 기술동향정보 메타데이터에 관한 연구", 한국문헌정보학회지, 39권3호.
- [2] 언론사 공동 이미지 아카이브 홈페이지 (<http://image.newsbank.co.kr>)
- [3] 한국정보사회진흥원(2005), "(개정판)알기 쉬운 UCI".
- [4] 한국정보사회진흥원(2006), "UCI 명세서 Version 2.1"
- [5] 한국언론재단(2007), "뉴스 전송 표준화 모델 개발 연구보고서"
- [6] IPTC 공식 홈페이지 ([www.iptc.org](http://www.iptc.org))
- [7] UCI 공식 홈페이지 ([www.uci.or.kr](http://www.uci.or.kr))