

뉴스 교환 포맷 표준 NewsML

한겨레신문(주) | 정상택

1. 머리말

전자책 63만 권을 제공하는 아마존의 킨들, 디지털 콘텐츠 산업 구조를 재편한 애플의 아이패드 등 새로운 디지털 기기가 보급되면서 디지털 콘텐츠의 생산과 소비가 날로 가속화하고 있으며, 기기와 콘텐츠의 포맷 표준을 선점해 시장을 차지하려는 업계의 경쟁도 치열하다. 국내에서도 이런 상황을 인식하고 유관 단체와 기업들이 한데 모여 ‘전자출판물 표준화 포럼’을 설립해 디지털 콘텐츠 표준화에 적극적으로 대응하고 있는 것은 다행스러운 일이다.

오래된 전통 매체인 뉴스 분야에서도 이러한 미디어 환경 변화에 상응하여 다양한 연구·개발과 디지털 콘텐츠 표준화를 진행하고 있다. 그 중 특히 IPTC¹⁾의 NewsML, NITF²⁾는 국·내외의 우수한 언론사에서 폭넓게 사용하는 뉴스교환포맷으로 무선인터넷의 확산과 스마트폰 같은 디지털 기기의 보급에 따라 점차 활성화될 것으로 보인다. 이런 추세에 맞춰 IPTC의 뉴스교환포맷 표준인 NewsML-G2를 이 글에서 간략하게 소개하고자 한다.

2. NewsML-G2

2.1 개요

뉴스는 배포되어야 하는 특성이 있다. 만일 뉴스 콘텐츠에 관한 표준이 부재하고 다양한 뉴스 콘텐츠 포맷이 난립한다면 뉴스 제공자와 소비자는 많은 불편을 감수해야 하고 불필요한 비용을 내야 할 것이다. 이런 문제를 해결하기 위해 IPTC는 NewsML 1.0부터

G2-표준에 이르기까지 뉴스 콘텐츠 교환과 관리를 위한 표준체계를 제정하게 되었다.

NewsML은 뉴스 교환을 목적으로 설계되었으나 현재는 뉴스의 취재, 편집, 배포, 아카이빙 등 뉴스 관련 공정의 전 영역에 두루 쓰이고 있다. 특히 NewsML-G2에서는 이전 모델보다 데이터 구조가 간결해졌고 시멘틱 웹과 지식기반 시스템을 구현할 수 있는 토대를 갖춘 것이 큰 특징이다.

NewsML은 XML을 기반으로 하며 뉴스의 내용뿐만 아니라 다양한 메타데이터, 뉴스 처리에 관한 지시, 뉴스 패키징에 관한 정보, 전송 정보 등 뉴스와 관련된 모든 요소와 속성을 정의·표현할 수 있다. XML을 기반으로 하지만 NITF와 같은 텍스트 기반의 마크업, XHTML, Atom³⁾과 RSS, hNews⁴⁾와 같은 마이크로포맷과도 호환할 수 있다.

국외에서는 AFP, AP, UPI, Dow Jones, Thomson Reuters, Wall Street Journal Online 등 30여 언론사와 20여 시스템 업체에서 사용 중이고, 국내에서는 경향신문, 세계일보, 조선일보, 한겨레 등 50여 언론사와 한국언론진흥재단 등에서 사용하고 있다.

2.2 데이터 모델 구조

NewsML-G2는 뉴스 아키텍처(NAR, News Architecture)로 불리는 공용 프레임워크를 바탕으로 한다. 뉴스 아키텍처는 네 개의 아이템 컨테이너 클래스와 이 클래스의 조합으로 구성된 상위 클래스 anyItem으로 구성된다. 네 개의 아이템 컨테이너 클래스는 각각 다음과 같은 정보를 포함한다(그림 1 참조).

1) International Press Telecommunications Council. 국제언론통신위원회. 세계 주요 통신사, 언론, 뉴스산업체 등 70여 단체가 구성한 컨소시엄으로 영국 런던에 있다. 뉴스교환에 관한 기술표준 개발과 관리, 보급에 관한 일을 한다. <http://www.iptc.org>

2) News Industry Text Format. 뉴스산업 텍스트 포맷. Associated Press, Dow Jones, The New York Times, Deutsche Presse-Agentur 등에서 사용하고 있다. 점차 NewsML로 옮겨가는 추세이다.

3) 웹로그, 최신 뉴스 등 웹 콘텐츠 신디케이션을 위한 XML 기반의 문서 포맷. <http://www.atomenabled.org/>

4) Associated Press와 Media Standards Trust가 hAtom을 확장하여 만든 마이크로 포맷. 뉴스 제작 및 배포에 필요한 다양한 필드 값을 도입, 다채로운 시멘틱 정보를 제공한다. <http://microformats.org/wiki/hnews>

표 1 NAR 클래스

newsItem	텍스트, 사진, 비디오 및 오디오 뉴스 콘텐츠와 메타데이터 http://www.iptc.org/std/NAR/1.5/specification/NAR_1.5-spec-Framework-Core.xsd
packageItem	뉴스 및 체계화된 데이터 참조로 뉴스 패키지를 구성 http://www.iptc.org/std/NAR/1.5/specification/NAR_1.5-spec-PackageItem-Core.xsd
conceptItem	인물, 기관, 장소와 같은 실재 정보와 제목, 주제와 같은 추상 정보 http://www.iptc.org/std/NAR/1.5/specification/NAR_1.5-spec-ConceptItem-Core.xsd
knowledgeItem	유관 콘셉트와 관련해 수집, 분류한 세부 정보 데이터베이스 http://www.iptc.org/std/NAR/1.5/specification/NAR_1.5-spec-KnowledgeItem-Core.xsd

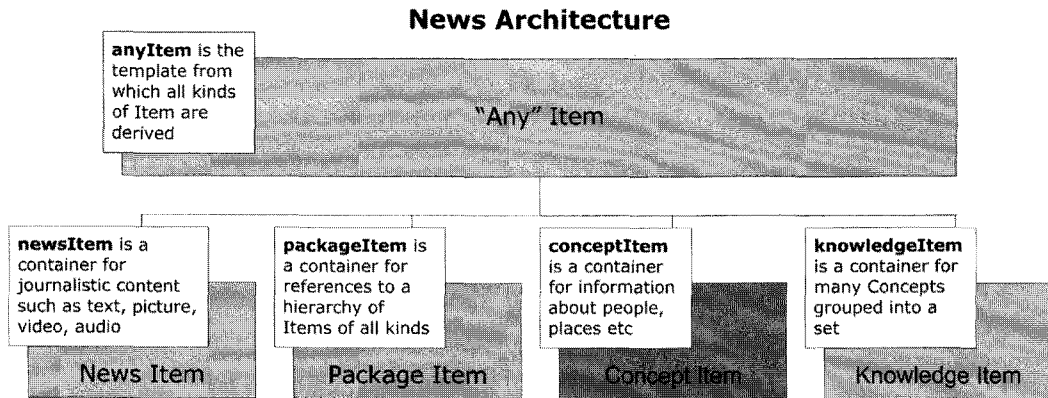


그림 1 네 개의 아이템을 유기적으로 관리할 수 있는 뉴스 아키텍처 구조

뉴스 아키텍처는 형식과 활용이 독립적이기 때문에 지향 소프트웨어로 개발할 수 있으며, 뉴스 아키텍처의
IPTC에서 제공하는 XML 스키마를 Java와 같은 객체 목적은 시멘틱 웹 환경에 적합하게 뉴스 콘텐츠를 구

표 2 NAR 스키마 분해

schema location:	http://www.iptc.org/std/NewsML-G2/2.4/specification/NAR_1.5-spec-Framework-Core.xsd
attribute form default:	unqualified
element form default:	qualified
targetNamespace:	http://iptc.org/std/nar/2006-10-01/
Elements	accountable, audience, br, broader, by, catalog, concept, contentCreated, contentModified, contributor, copyrightHolder, copyrightNotice, created, creator, creditline, dateline, definition, description, edNote, embargoed, facet, fileName, firstCreated, genre, geoAreaDetails, headline, icon, infoSource, itemClass, keyword, language, located, name, narrower, note, objectDetails, organisationDetails, personDetails, POIDetails, provider, pubStatus, related, remoteInfo, role, sameAs, service, slugline, subject, title, urgency, usageTerms, versionCreated
Groups	AdministrativeMetadataGroup, ConceptRelationshipsGroup, DescriptiveMetadataGroup, EntityDetailsGroup, ItemManagementGroup, RecurrenceGroup
Complex types	AnyItemType, ApproximateDateTimePropType, BlockType, ConceptIdType, ContactInfoType, ContentMetadataType, DateOptTimePropType, DateTimeOrNullPropType, DateTimePropType, ElectronicAddressType, FlexPropType, GeoCoordinatesType, IntlStringType, ItemMetadataType, LabelType, LinkType, QCodePropType, QualPropType, RightsInfoType, TruncatedDateTimePropType, TypedQualPropType,
Simple types	ByDayListType, ByHourListType, ByMinuteListType, ByMonthDayListType, ByMonthListType, BySecondListType, BySetposListType, ByWeekNoListType, ByYearDayListType, DateOptTimeType, EmptyStringType, Int100Type, Int1to9Type, IRIType, QCodeType, TruncatedDateTimeType, UnionDateTimeEmptyStringType, UnionDateTimeType
Attr. groups	deprecatedLinkAttributes, flexAttributes, i18nAttributes, rankingAttributes, recurrenceRuleAttributes, targetResourceAttributes

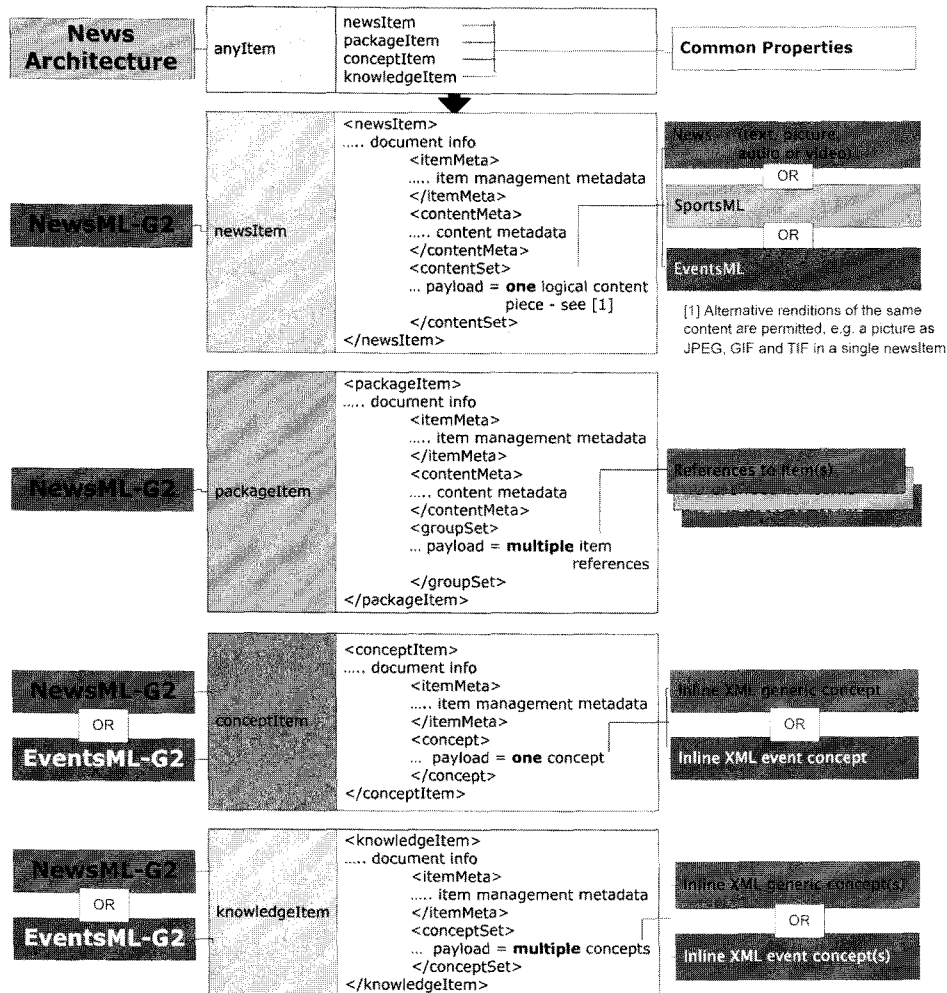


그림 2 뉴스 아키텍처와 G2-표준의 활용 사례

조화하여 관리하고, 다른 지식기반 솔루션과 연계하는 것이다.

뉴스 아키텍처 프레임워크를 바탕으로 멀티미디어를 포함한 모든 유형의 뉴스 콘텐츠를 관리할 수 있는 NewsML-G2, 날짜순으로 제공되는 뉴스 아이템이나 대형사건 관련한 뉴스 콘텐츠를 입체적으로 제공하는 EventML-G2, 일반 스포츠 경기의 기록, 선수 이력 등 스포츠 분야의 구체적 통계를 관리할 수 있는 SportsML-G2가 있다. 이 세 가지 모델을 한데 묶어 'G2-표준'으로 부른다(그림 2 참조).

2.3 메타데이터

G2 표준의 원칙은 콘텐츠와 메타데이터를 분리하는 것이다. 텍스트 위주의 뉴스 모델에서는 메타데이터의 종류가 많지 않았으나 현재의 멀티미디어 환경에서는 사정이 다르다. 웹만 보더라도 지면에선 존재하지 않던 페이지 URL, 유관 콘텐츠와의 연결 링크, 미디어 유형 등 콘텐츠의 특성을 설명하고 관리할 많은

메타데이터가 생겨났다.

게다가 콘텐츠의 생산량이 급속하게 증가하기 때문에 사용자들은 자신이 원하는 정보를 찾는 데 더 많은 노력과 비용을 소비해야 한다. 이를 해결하는 방안 중 하나는 각 콘텐츠와 관련된 메타데이터를 속성별로 더욱 정교하게 분류하고 이를 서로 교환하거나 공유하는 것이다.

2.3.1 G2-표준 속성

G2-표준은 표준의 유지, 확장을 간편하게 해 일관

표 3 G2-표준의 속성들

① String, Block & Label Type	일반 텍스트 사용
② Qualified Properties	미리 정의한 제한된 어휘만 Q코드 형식으로 사용
③ Flexible Properties	각 속성을 자유롭게 선택 사용
④ Other Property Types	날짜, 시간, 숫자, 링크
⑤ Property Groups	elements & attribute group

성을 유지하고 각 속성을 일반화하여 실제 업무에 쉽고 편리하게 적용할 수 있도록 한 것이다. G2-표준의 각 속성은 재사용이 가능한 서식을 사용하며 다음 중 하나 또는 몇 가지를 혼합해 사용한다.

2.3.2 메타데이터 통제

메타데이터에 특정한 조건을 부가해 메타데이터를 통제할 수 있다. 예를 들면, '사진', '포토', 'photograph', 'photo', 'picture', 'pic'은 모두 사진을 의미한다. 사람은 이 단어들을 사진으로 인식할 수 있으나 시스템에서는 특정한 값을 지정하지 않으면 처리가 곤란해진다. 그래서 별도의 코드스킴이나 분류를 통해 특정한 값을 지정해 주어야 하는데 이를 '통제어휘(CV, Controlled Values)'라 부르며, Q코드(QCodes)를 통하여 필요한 메타데이터 접근할 수 있다.

또 IPTC에서는 뉴스 교환에 필요한 다양한 CV 세트를 마련해 놓고 있는데 이를 '뉴스코드(NewsCodes)'로 부른다.

2.4 적합성 모델

각 아이템은 아이템 간에 연결된 개념과 프로세스의 복잡성에 따라 'Core Conformance Level(CCL)'과 'Power Conformance Level(PCL)' 모델로 구분한다. CCL은 간결함과 상호운용성 측면에 초점을 맞춘 것이고, PCL은 CCL을 포함하는 슈퍼세트로 사용자가 좀 더 자유롭게 응용할 수 있도록 확장한 모델이다.

다만, PCL을 사용하는 경우는 송신시스템과 수신시스템이 일치해야 하므로 프로세스의 복잡성이 증가하고, 상호운용성이 낮아지는 단점이 있다. 예를 들어, 송신 측에서 PCL을 사용하고 수신 측에서 CCL을 사용한다면, 수신 측에서는 CCL로부터 확장한 송신 측의 PCL 문서를 해석할 수 없다. 따라서 사용자는 아이템 모델('core' 또는 'power')을 반드시 문서 안에 명기해 상호 간에 인식할 수 있도록 해야 한다.

2.5 버전

2000년 10월, NewsML 1.0이 최초 승인되었고, 2002년 10월에 NewsML 1.1, 2003년 10월에 NewsML 1.2가 승인되었다. 이후 사용자들의 경험을 평가해 '모든 미디어에서, 모든 유형의 데이터를, 효율적이고, 신뢰할 수 있고, 포괄적이며 공동으로 사용할 수 있는' 방식으로 보완해 2004년 G2-표준을 발표하였다.

NewsML 1.x는 멀티미디어 뉴스와 뉴스 패키지 교환에 초점을 두었지만, G2-표준은 웹 2.0, 시멘틱 웹 등 새로운 기술에 상응하기 위해 새로 구성한 것이다.

3. 기본 문서 예시

IPTC나 로이터 웹사이트에서는 멀티미디어, 뉴스 패키징 등 다양한 형태의 NewsML 샘플 문서를 제공한다. 여기서는 단순히 `newsItem` 하나로만 구성된 기본적인 NewsML-G2 문서를 분석하면서 주요 속성과 요소들을 살펴본다.

NewsML-G2에서 `newsItem`은 기사일 때 그림 3과 같이 네 부분으로 구성된다.

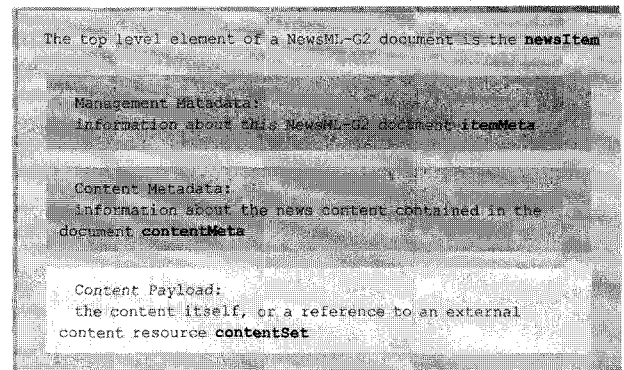


그림 3 NewsML-G2 뉴스 아키텍처로 구성한 뉴스 아이템

- 최상위 레벨 요소 `<newsItem>`은 문서의 고유 id (GUID), XML 네임스페이스, 카탈로그 참조에 관한 저작권과 라이선스 정보를 포함한다. 이 요소는 다른 세 개의 아이템에서도 공통적인 형식으로 사용되기 때문에 'AnyItem Type'으로 정의되어 있다.
- 관리 메타데이터(Management Metadata). `<itemMeta>`는 아이템클래스와 같은 콘텐츠를 관리하는 데 필요한 속성과 NewsML-G2 문서에 관한 정보를 제공한다.
- 콘텐츠 메타데이터(Content Metadata). `<contentMeta>`는 뉴스 카테고리나 같은 설명 메타데이터(descriptive metadata)와 정보출처, 타임스탬프와 같은 관리 메타데이터(administrative metadata)와 콘텐츠 정보를 포함한다.
- 콘텐츠(Content)는 `<contentSet>` 포장(wrapper) 속에 포함된 모든 미디어 유형을 포함한다. 텍스트 기사와 같은 문자 콘텐츠는 XHTML, NITF처럼 인라인(in-line)의 형태로 선택한 포맷 속에 포함될 수 있으며, 크기가 작은 바이너리 객체도 인라인의 형태로 포함될 수 있다. 사진, 비디오, 그래픽과 같은 바이너리 콘텐츠는 주로 외부 자원의 참조로 포함한다.

아래의 <리스트 1>은 앞으로 살펴볼 NewsML-G2 문서 전문으로 각 요소를 그림 3에 표시된 색상과 같게 `newsItem`, `itemMeta`, `contentMeta`, `contentSet` 순으로 표시하였다.

<리스트 1> Example of a NewsML-G2 Document

```
<?xml version="1.0" encoding="UTF-8" ?>
<newsItem guid="urn:newsml:iptc.org:20091107:tutorial-text-xhtml"
version="1"
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
NewsML-G2_2.4-spec-NewsItem-Core.xsd"
  standard="NewsML-G2"
  standardversion="2.4"
  xml:lang="en-GB" >
  <catalogRef
href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_11.xml" />
  <itemMeta>
  <itemClass qcode="ninar:text" />
  <provider literal="IPTC" />
  <versionCreated>2009-11-07T12:12:12+01:00</versionCreated>
  <pubstatus qcode="stat:usable" />
  </itemMeta>
  <contentMeta>
  <contentCreated>2009-11-06</contentCreated>
  <located literal="Washington" />
  <creator literal="JLaMeur">
  <name>Laurent J LaMeur</name>
  </creator>
  <language tag="en-GB" />
  <subject type="cpnat:abstract" qcode="subj:01000000" />
  <subject type="cpnat:abstract" qcode="subj:01026000" />
  <subject type="cpnat:abstract" qcode="subj:01026002">
  <name>art, culture and entertainment &qt; news &qt; media</name>
  </subject>
  <subject type="cpnat:organisation" literal="IPTC" />
  <slugline>IPTC-NewsML G2</slugline>
  <headline>NewsML-G2 is approved</headline>
  </contentMeta>
  <contentSet>
  <inlineXML contenttype="application/xhtml+xml">
  <html xmlns="http://www.w3.org/1999/xhtml">
  <head>
  <title>IPTC-NewsML G2</title>
```

```

</head>
<body>
  <h1>NewsML-G2 version 2.4 approved by the IPTC in Washington</h1>
  <p>
    <span class="dateline">Washington, November 6 (IPTC) </span>
    The International Press Telecommunications Council Standards
    Committee
    today approved the latest version of its News Exchange Format
    NewsML-G2.
  </p>
</body>
</html>
</inlineXML>
</contentSet>
</newsItem>

```

3.1 최상위 레벨 <newsItem>

첫 번째 줄은 XML 표준 문서 표시이다.

```
<?xml version="1.0" encoding="UTF-8" ?>
```

두 번째 줄에 있는 최상위 레벨 <newsItem>은 필수적으로 지정해야 하는 속성, NewsML-G2 문서의 GUID, 버전을 지정한다.

```
guid="urn:newsml:iptc.org:20081007:tutorial-text-xhtml" version="1"
```

GUID는 전 세계에서 유일한 id라야 한다. 예시에서 사용한 속성은 IPTC에 등록된 URN을 사용한다. 그러나 뉴스 제공자들은 고유의 스킴을 사용할 수도 있다. 제공자들이 주로 선택하는 스킴은 tag URI이다.

IPTC 네임스페이스와 XML 스키마에 대한 W3C 네임스페이스는 다음과 같다.

```
xmlns="http://iptc.org/std/nar/2006-10-01/
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
```

3.1.1 스키마 참조

스키마 참조는 로컬 파일시스템상에서 NewsML-G2 2.4 버전의 IPTC 스키마에 있으며 IPTC 네임스페이스와 연관이 있다.

```
xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
NewsML-G2_2.4-spec-NewsItem-Core.xsd"
```

스키마 참조는 중요한 부분으로, 사용하는 미디어 시스템이 문서에 있는 내용대로 작동하고 처리할 수 있는지 유효성을 검증해야 하며, 가능한 수행 수준과 서비스 관계도 참작해야 한다. 문서의 유효성을 검증할 때는 반드시 해당 지역에 저장된 IPTC G2 스키마와 비교하여 검증해야 한다.

3.1.2 G2 표준과 버전

표준 및 표준 버전은 IPTC G2-표준 스키마와 일치해야 한다.

```
standard="NewsML-G2" standardversion="2.4"
```

@standard에 사용할 수 있는 문자열은 'NewsML-G2' 또는 'EventsML-G2'뿐이다. 뉴스 아키텍처 버전에 따라 사용 가능한 @standard, @standardversion은 다음과 같다.

표 4 G2 표준 및 버전

NAR 버전	@standard	@standardversion
1.1	NewsML-G2	2.0
	EventML-G2	1.0
1.2	NewsML-G2	2.1
1.3	NewsML-G2	2.2
	EventML-G2	1.1
1.4	NewsML-G2	2.3
	EventML-G2	1.2
1.5	NewsML-G2	2.4
	EventML-G2	1.3

대규모의 시스템에서 IPTC의 스키마를 직접 참조하는 것은 시스템의 성능과 서비스에 많은 영향을 미치기 때문에 문서 내에 위 속성을 명시하여 해당 XML 스키마를 알려주는 편이 좋다.

3.1.3 적합성 모델

만약, 해당 문서가 PCL 모델을 사용한다면 반드시 그 사실을 명시해야 한다. PCL 사용 여부를 명시하지 않으면 CCL 모델을 사용한 것으로 간주한다.

```
conformance="power"
```

3.1.4 카탈로그 참조

```
<catalogRef href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_11.xml"/>
```

필요한 IPTC의 뉴스코드 카탈로그를 식별해 G2 프로세스가 Q코드를 역 참조할 수 있도록 한다. 문서에서 사용된 모든 카탈로그는 이 부분에 명시되어야 한다.

IPTC는 모든 카탈로그를 캐시 시스템에 저장하는 것을 추천한다. 카탈로그 제공자의 서버가 서비스 요청을 처리할 때 시스템에 걸리는 부하가 작고 뉴스 작업 공정에서도 부담이 적기 때문이다.

문서에 사용되는 기본 언어를 영국식 영어로 명시할 수도 있다. 그러나 문서 안의 XML에만 적용할 수 있다. 콘텐츠 내용은 언어에 개의치 않으며 콘텐츠 메타데이터에서는 다른 속성을 사용한다.

```
xml:lang="en-GB"
```

3.2 아이템 메타데이터 <itemMeta>

아이템 메타데이터는 <newsItem>에 관한 정보와 처리방식을 알려준다. <itemClass>, <provider>, <versionCreated>로 구성되며 반드시 명시해야 한다.

예시의 첫 번째 줄에 있는 qcode는 수신 측에 IPTC의 뉴스코드 스킴을 사용하여 문서의 콘텐츠가 텍스트임을 알려준다.

```
<itemClass qcode="ninat:text" />
```

provider는 뉴스 제공자에 관한 정보이다.

```
<provider literal="IPTC" />
```

<provider> 요소는 속성 지정에 유연성이 있어 Q코드나 문자열 모두 사용할 수 있다.

versionCreated는 뉴스 아이템(콘텐츠가 아니다.)이 작성된 일시를 UTC의 오프셋을 포함하여 표시한다. ISO 8601 형식을 사용한다. YYYY-MM-DDThh:mm:ss±hh:mm.

```
<versionCreated>
```

```
2008-10-07T12:12:12+01:00
```

```
</versionCreated>
```

<pubStatus> 속성은 뉴스의 발행 상태를 표시한다. 속성값이 'usable'이면 이 속성을 생략할 수 있다. 만약 <putStatus>가 생략되었다면 기사와 해당 콘텐츠를 사용할 수 있는 것으로 간주한다. 아래처럼 표기한다.

```
<pubStatus qcode="stat:usable" />
```

pubStatus는 G2의 뉴스 제공자들, 특히 뉴스에이전 시에는 매우 중요한 속성이기 때문에 반드시 명시해야 하며, IPTC의 발행 상태를 표시하는 뉴스코드는 필수 요소이다. 이 뉴스코드를 나타내는 별칭은 'stat'이며, 이 별칭이 표시되면 속성값은 'usable'로 세팅된다. 스킴이 허용하는 또 다른 값들은 다음과 같다.

- canceled: 뉴스 콘텐츠 또는 참조 모두 사용할 수 없다.
- withheld: 추가 공지가 내려지기 전까지 콘텐츠를 사용할 수 없다.

취소와 보류의 전형적인 예는, 뉴스 제공자가 해당 콘텐츠를 이미 배포했으나 법적인 이유로 취소 또는 보류할 필요가 생길 때이다. 기사가 취소되면 사용할 수 없으나 보류되면 다음에 'usable'로 변경될 수 있다.

3.3 콘텐츠 메타데이터 <contentMeta>

콘텐츠 메타데이터는 G2 아이템 콘텐츠와 관련되며 아래와 같이 두 종류가 있다.

- 관리 메타데이터(Administrative Metadata): 정보의 출처, 긴급성과 같은 정보
- 설명 메타데이터(Descriptive Metadata): G2 아이템의 주제, 즉 카테고리, 헤드라인과 같은 정보

필수적으로 명시해야 하는 요소는 없으나 본 예시에는 대다수 뉴스 제공자들이 주로 사용하는 요소들이 표기되어 있다.

<contentCreated> 속성은 날짜로 YYYY-MM-DD 포맷을 사용하며 선택적으로 시간과 시간대를 포함할 수 있다. 이 속성은 날짜를 간략하게 표현할 수 있는데 우측부터 좌측으로 가면서 끝 부분을 생략할 수 있다. 단, YYYY는 반드시 표기해야 한다.

```
<contentCreated>2009-11-06</contentCreated>
```

<located> 요소는 콘텐츠가 만들어진 장소에 대한

정보로 반드시 사건이 발생한 장소여야 한다. 통제어휘 대신에, "Washington"처럼 지역명을 그대로 쓴다.

```
<located literal="Washington" />
```

콘텐츠의 작성자에는 문자열 또는 통제어휘를 사용한다.

```
<creator literal="LLeMeur">
  <name>Laurent Le Meur</name>
</creator>
```

콘텐츠 언어는 IETF's BCP 47을 사용하여 표기한다.

```
<language tag="en-GB" />
```

주제코드는 뉴스 수신자들에게 콘텐츠의 카테고리를 알려준다. 표기방법은 IPTC의 뉴스코드 표기법을 따른다.

```
<subject type="cpnat:abstract" qcode="
  "subj:01000000" />
<subject type="cpnat:abstract" qcode="
  "subj:01026000" />
<subject type="cpnat:abstract" qcode="
  "subj:01026002">
  <name>art, culture and entertainment
    &gt; news &gt; media</name>
</subject>
```

예시에서 사용된 주제 요소는 두 가지 속성, 즉 type과 qcode를 갖는다. 이 두 가지 속성 모두 Q코드로 표기한다. 첫 번째 'cpnat:abstract'는 IPTC 뉴스코드 스킴을 참조하여 Q코드 'subj:01000000'로 표시한다. 여기서는 Q코드의 내용이 인물, 조직과 같은 실제 정보가 아니라 추상적 개념임을 알 수 있다.

이는 G2-표준만이 가지는 중요한 특징이다. 예를 들어, 애플리케이션은 문서가 참조하는 인물과 관련된 개념을 모두 검색하여 메타데이터로 코드화되어 있는 뉴스기사, 관련된 인물 등을 찾을 수 있다. 검색 후에는 자동으로 관련 자료와 링크를 걸 수도 있다.

예시 문서의 다음 주제 유형은 조직이며 Q코드로 표기하지 않고 문자열로 직접 표현하였다.

```
<subject type="cpnat:organisation"
  literal="IPTC" />
```

아래는 콘텐츠의 내용을 비공식적으로 알려주려고 콘텐츠 제공자들이 자주 사용하는 슬러그라인이다.

```
<slugline>IPTC-NewsML G2</slugline>
```

그리고 <contentMeta> 부분의 마지막 줄은 헤드라인 요소에 관한 것이다.

```
<headline>NewsML-G2 v2.4 is approved</
  headline>
```

콘텐츠 자체 내에 헤드라인이 포함될 수도 있다. 기사 내에 헤드라인을 포함할지를 결정하는 것은 뉴스 제공자의 몫이다. <contentMeta> 헤드라인은 콘텐츠와 별개로 사용할 수 있으나 바이너리 콘텐츠를 사용할 때는 원문을 식별할 방안이 있어야 한다.

3.4 콘텐츠 (contentSet)

본 예시에서 사용한 inline XML 텍스트를 표시한다.

```
<inlineXML contenttype="application/
  xhtml+xml">
```

4. 맺음말

이상으로 NewsML-G2를 간략하게 살펴보았다. 더 자세한 내용은 참고문헌이나 IPTC에서 제공하는 정보를 참고하시기 바란다.

현재 뉴스 콘텐츠 시장의 흐름을 보면 텍스트 콘텐츠가 점차 감소하고 비디오와 같은 멀티미디어 콘텐츠가 증가하고 있다. 뉴스 유통 부문에서도 전통적인 B2B 방식에서 점차 다양한 소비층으로 공급되는 이른바 B2X 형태가 대세이다. 즉, 멀티미디어, 멀티채널, 멀티플랫폼 환경으로 변하지 않고서는 경쟁 무대에 설 수 없다는 의미이다. 또한, 끊임없이 내려오는 비용 절감에 대한 압력, 새로운 서비스와 비즈니스 모델에 대한 개발 요구, IT 투자에 대한 심각한 통제, 급변하는 기술과 높아진 뉴스 소비자의 욕구 충족 등도 뉴스 산업 참여자가 풀어야 할 숙제들이다.

IPTC의 G2-표준은 이러한 미디어 환경 변화의 중심에서 있는 유수의 언론사 및 관련 시스템 업체들이 오랜 기간 연구하며 개발, 보완한 결과물이다. 한 번쯤 시간을 내어 내용을 훑어보는 것도 유익할 듯싶다.

참고문헌

- [1] NewsML-G2 명세 관련 문서들:
http://www.iptc.org/std/NewsML-G2/2.4/specification/NAR_1.5-spec-Framework-Core.xsd,
http://www.iptc.org/std/NewsML-G2/2.4/specification/NewsML-G2_2.4-spec-NewsItem-Core.xsd,
http://www.iptc.org/std/NewsML-G2/2.4/specification/NewsML-G2_2.4-spec-CCL.pdf,
<http://www.iptc.org/std/NewsML-G2/2.1/specification>

/schema-doc-power/NewsItem.html

- [2] NewsML-G2 Introduction,
http://iptc.org/site/News_Exchange_Formats/NewsML-G2/Introduction/
- [3] Kelvin Holland, IPTC-G2-Implementation-Guide_2.pdf,
p.16, p.23-32, IPTC, 2009.
- [4] A point by point account of the benefits of NewsML,
<http://about.reuters.com/newsml/whynewsml-keybenefits.asp>

약 력



정상택

2000 정보통신연구원, 소프트웨어 개발실장 겸
리눅스매거진 편집장

2010 한겨레신문(주), 정보기술센터장

관심분야 : - Software Programming Language
- Mobile Web & Application Development
- Human Interface & Universal Design

E-mail : windst@hani.co.kr