

Using MPEG-21 to repurpose, distribute and protect news/NewsML information

Pierfrancesco Bellini, Ivan Bruno, Paolo Nesi

DISIT-DSI, Distributed Systems and Internet Technology Lab

Dipartimento di Sistemi e Informatica, Università degli Studi di Firenze, Firenze, Italy

<http://www.disit.dsi.unifi.it/>, nesi@dsi.unifi.it, paolo.nesi@unifi.it

Abstract

The distribution of news is a very articulated and diffuse practice. To this end one of the most diffuse formats for news production and distribution is the NewsML. The management of news has some peculiarities that could be satisfied by using MPEG-21 as container and related production tools and players. To this end, an analysis of modeling NewsML with MPEG-21 has been performed and reported in this table. The work has been performed for AXMEDIS project which is a large IST Research and Development Integrated Project of the European Commission.

1. Introduction

At present, there is a large number of content formats ranging from the simple files: documents, video, images, audio, multimedia, etc., to integrated content models for packaging such as MPEG-21 [1], [5], SCORM, MXF, NewsML [6], SportML, etc. These models try to wrap/refer digital resource/essences and in some cases to wrap them in a digital container, so as to make them ready and simpler for delivering. Among these formats the ones used for distributing and sharing news are mainly text and XML oriented such as NewsML of IPTC (International Press Tele-communication Council). Recently a new version of NewsML has been proposed, the NewsML-G2 that provides support for referencing textual news, resource files, for paging them, while collecting metadata and descriptors, vocabularies, etc. (<http://www.iptc.org>). Furthermore, the news are typically massively processed by news agencies and/or by TV news redactions. They are not only received in NewsML formats but also in HTML, plain TXT, PDF formats as well. The agencies and redactions need to move, transcode, and adapt them to different formats processing both text and digital essences, by changing resolution, summarizing text, adapting descriptive metadata, etc. In some cases, the adaptation has to be performed on demand as a result of an answer to a query or request to a database or on a web service.

Moreover, frequently the news contain videos and images, while solution proposed by NewsML of zipping the file constrains the users to unzip the files in some directory to access and play the video. In addition, news contain frequently sensitive data for which protection of IPR (intellectual property rights) is needed. Thus, most of the above mentioned formats present a number of problems such as limitations related to the adopted packaging format. For example, such as the NewsML limitations on the packaging so as to prevent from playing effectively video content from the package without decompressing and/or unpacking, and limitations on the protection and preservation of the IPR (intellectual property rights). Such problems are related to the file format and protection support including certification, content signature and licensing.

Among the formats mentioned, the AXMEDIS implementation of the MPEG-21 file format and MXF supports the direct play. Only the MPEG-21 also support a range of business and transaction models via a DRM (Digital Rights Management) solution and with a set of technological protection supports.

In this paper, a solution to solve the above mentioned problems of news modeling, massive production and processing and distribution is presented. The solution proposed is based on AXMEDIS content model and processing GRID platform, AXCP. AXCP provides a set of technical solutions and tools to allow the automation of cross media content processing, production, packaging, protection and distribution. AXMEDIS multimedia processing can cope with a large number of formats including MPEG-21 and it can work with a multichannel architecture for the production of content on demand [3]. AXMEDIS is a framework that has been funded by the European Commission and it has been developed by many partners including: University of Florence, HP, EUTELSAT, TISCALI, EPFL, FHGIGD, BBC, AFI, University Pompeo Fabra, University of Leeds, STRATEGICA, EXITECH, XIM, University of Reading, etc. More

technical information, as well as how to make registration and affiliation to AXMEDIS can be recovered on <http://www.axmedis.org>

In order to solve the above described problems, the AXCP solution has been augmented by semantic processing capabilities, NewsML modeling and conversion strategy into AXMEDIS MPEG-21 format with the aim of preserving semantics and capabilities of the early news files processed [4], [5]. In this case, the MPEG-21 models and tools have been used: (i) as a descriptor and/or a container (with AXMEDIS file format) of information and multiple file formats, (ii) as a vehicle to protect the IPR when the information is distributed towards non protected channels or it contains sensitive information.

The paper is organized as follows. In section 2, a short overview of AXMEDIS content processing platform for multimedia processing is reported. Section 3 refers to modeling of NewsML into MPEG-21 and AXMEDIS formats. In Section 4, some implementation details regarding the AXCP are reported. An analysis of the advantages identified in using the AXMEDIS model and tools are reported in Section 5. Conclusions are drawn in Section 6.

2. AXMEDIS Content Processing

The AXCP tool is based on a GRID infrastructure constituted of a Rule Scheduler and several Executors for process executing. AXCP Rules are formalized in AXCP java script [2], [4]. The AXCP Rule Scheduler performs the rule firing, discovering Executors and managing possible problems. The scheduler may receive commands (to invoke a specific rule with some parameters) and provide reporting information (e.g. notifications, exceptions, logs, etc...) to external workflow and tools by means of a WEB service.

The Rule Executor receives the Rules to be executed from the Scheduler and performs the initialization and the launch of the Rule. During the run, the Executor could send notifications, errors and output messages to the Scheduler. Furthermore, the Executor could invoke the execution of other Rules sending a specific request to the Scheduler, in order to divide a complex Rule/procedure into sub rules/procedure running in parallel, thus allowing a rational use of the computational resources accessible in the content factory, on the GRID. This solution maintains advantages of a unified solution and allows enhancing the capabilities and the scalability of the AXMEDIS Content Processing.

The AXCP processing tools are supported by a Plugin technology which allows each AXCP Rule Executor to link dynamically any content processing

tool and algorithm (e.g. audio, video and image adaptation, transcoding, encryption) and to cope with possible customized algorithms and tools.

As to the processing capabilities, an AXCP Rule formalises in its own language features to perform activities of ingestion, query and retrieval, storage, adaptation, extraction of descriptors, transcoding, synchronisation, fingerprint, indexing, summarization, metadata manipulation and mapping via XSLT, packaging, protection and licensing in MPEG-21 and OMA, publication and distribution via traditional channels and P2P.

3. From NewsML to AXMEDIS modeling passing via MPEG-21

The NewsML has a structure at 4 nested levels (from the contained to the smaller components): *NewsEnvelope*, *NewsItem*, *NewsComponent* e *ContentItem* (<http://www.iptc.org>).

The *News Component* mainly contains the information that may be used for modeling the *NewsItems*. At the end the *ContentItem* describes the contribution in terms of comments, classification, media type, format, notation, etc. The NewsML has also metadata mapped in the architecture and in particular in the *NewsComponent*: Administrative Metadata, Descriptive Metadata, and Rights Metadata. The information for the news identification are reported into the *NewsItems*, each of them can be univocally identified.

On the basis of our analysis, we have identified 6 main entities which have to be addressed: NewsML, NewsEnvelope, NewsItem, NewsComponent, ContentItem, TopicSet, Catalog (see Figure 1)

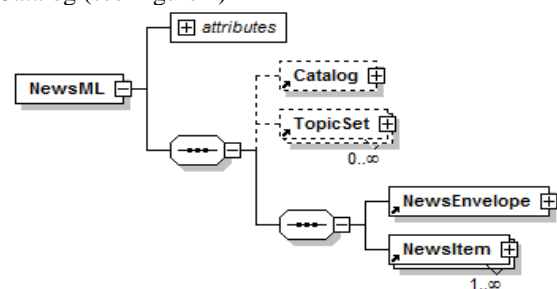


Figure 1 – NewsML main entities

The resulting model is hierarchical and in order to be ingested, analyzed and converted it has been replicated into an object oriented model allowing us to represent this model in the memory, by considering their

relationships and roles, as in the UML diagram reported in Figure 2.

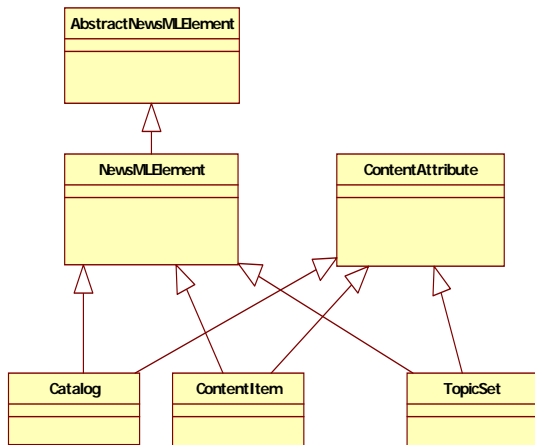


Figure 2 – Modeling NewsML main entities for conversion and analysis

In addition, also other classes have been implemented to model the NewsML such as: Topic, NewsMLDocument, NewsComponent, NewsItem also specialised from both NewsMLElements and ContentAttribute. The proposed model allows to ingest quickly the NewsML structures.

The realized model allows to perform the needed transformations on the NewsML files in an efficient manner. For example, the extraction of a NewsComponents removing its parts from the tree, the addition of news, etc, together with the conversions of the NewsML in other formats such as XML, HTML, Text and files, and MPEG-21 as described in the following.

The resulted model has been also analyzed to map the information into the MPEG-21 structure of the DIDL (Digital Item Description Language).

NewsML element	AXMEDIS Element	MPEG-21
Metadata	AXInfo + Dublin Core	Descriptor
NewsML	AxObject	Item
NewsItem	AxObject	Item
NewsComponent	AxResource	Component
ContentItem	AxResource	Component

Table 1 – Mapping concepts of NewsML to AXMEDIS and MPEG-21

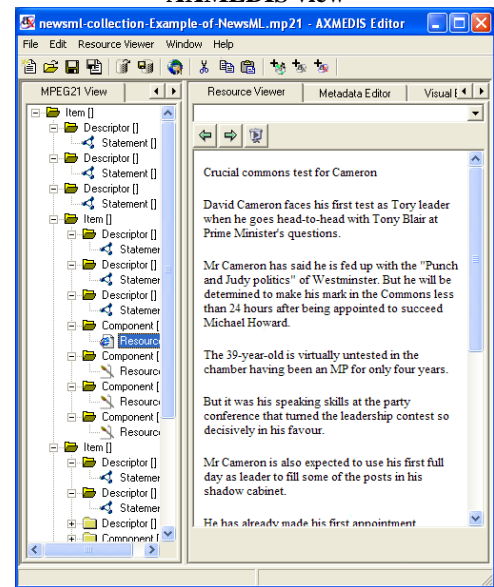
In Table 1, a mapping of the NewsML elements with those of MPEG-21 and AXMEDIS is provided.

The AXMEDIS editor allows you to see both MPEG-21 and AXMEDIS views of the newsML file as

depicted in Figure 3. AXMEDIS view is only a more abstract view of the AXMEDIS file format ISOMEDIA based. The AXMEDIS mapping is more effective and easy to understand than the underlining MPEG-21 modeling that is fully flat and hard to be understood by humans. The resulted MPEG-21 container of the News can be protected by using the MPEG-21 REL and AXMEDIS tools for DRM.



AXMEDIS view



MPEG-21 view

Fig.3 – A NewsML on the AXMEDIS Editor

In Figure 3, AXMEDIS view, the nesting levels of AXMEDIS objects are evident. They can be moved or

extracted simply using drag and drop. The same approach can be adopted to work with single contributions: text and/or digital files (images, video, etc.). They can be played directly into the editor and into the AXMEDIS player. An additional feature is the index in HTML of the converted NewsML items. It has been automatically produced by processing the NewsML structure in the AXCP script. That index is an HTML file enforced into the AXMEDIS Object (see the bottom of the tree in Figure 4).

4. Implementation on the AXCP GRID

The above mentioned object oriented module for NewsML ingestion, modelling and processing has been added to the AXCP Node engine. Therefore, a set of functionalities, API, to access the NewsML models has been defined and made directly accessible into the AXCP Java Script Multimedia processing language.

5. Benefits and results

This solution based on AXCP allowed to set up flexible automatic processes where NewsML information is ingested and processed in a very efficient manner, while considering any kind of conditions and structures for repurposing them and adapting news including text and digital essences towards different formats: HTML, TXT, PDF, MPEG-21, SMIL, etc., either integrating or not digital essences into them and distributing them via email, posting on FTP, on DBs, etc.

Besides, the news modeling with AXMEDIS has some advantages, as the resulting AXMEDIS object can be:

- used as a news descriptor and/or a news container (with AXMEDIS file format), supporting any kind of file formats for the digital essences being integrated into the news.
- used to manipulate the news, to add other information via AXMEDIS Editor and to make a directly play of the essences into the news without extracting them from the package.
- searched into the internal body of the news object, thus making the understanding and browsing of complex news easier, by adding simple Intelligent methods such as the ones described into [5].
- annotated conformant to MPEG-21 as described in [5].
- IPR protected when the information is distributed towards non protected channels or it contains sensitive information.
- distributed in several manners and accessed via PC, PDA, etc.

6. Conclusions

In this paper, the analysis of the modelling NewsML and news in general with MPEG-21 has been performed and presented. The results demonstrated that the structure of the News can be quite easily modelled in MPEG-21. In addition, the news processing consisting in their ingestion and transcoding can be performed on the AXCP platform in quite easy manner since now an ingestion module of NewsML has been developed and added. As a result, a number of advantages have been identified and demonstrated, as reported in Section 5. The full documentation can be recovered on the AXMEDIS portal <http://www.axmedis.org>. AXMEDIS is an open platform, which means that you can join the AXMEDIS community. The example mentioned in this paper is accessible from the same web portal.

Acknowledgments

The authors would like to express their thanks to all the AXMEDIS project partners including the Expert User Group and all the affiliated members, for their contribution, funding and collaboration efforts. A specific acknowledgment to EC IST for partially funding the AXMEDIS project. A warm thanks to all the AXMEDIS people who have helped us in starting up the project and sorry if they might have been not involved in the paper nor mentioned. We trust in their understanding.

References

- [1] MPEG Group, "Introducing MPEG-21 DID", <http://www.chiariglione.org/mpeg/technologies/mp21-did/index.htm>
- [2] J. Thiele, "Embedding SpiderMonkey - best practice" <http://egachine.berlios.de/embedding-sm-best-practice/embedding-sm-best-practice-index.html>
- [3] P. Bellini, I. Bruno, P. Nesi, "A language and architecture for automating multimedia content production on grid", Proc. of the IEEE International Conference on Multimedia & Expo (ICME 2006), IEEE Press, Toronto, Canada, 9-12 July, 2006.
- [4] P. Bellini, P. Nesi, D. Rogai, "Exploiting MPEG-21 File Format for cross media content", Proc. Of the International Conference on Distributed Multimedia Systems, DMS 2007, September 6-8, 2007, San Francisco Bay, USA, Organised by Knowledge Systems Institute.
- [5] P. Bellini, I. Bruno, P. Nesi, M. Spighi, "Intelligent Content Model based on MPEG-21", in Proc. AXMEDIS 2008, Florence, Italy, 17-19 Nov. 2008, pp 41-48, IEEE Press.
- [6] Kodama, M.; Ozono, T.; Shintani, T.; Aosaki, Y.; "Realizing a News Value Markup Language for News Management Systems Using NewsML" Complex, Intelligent and Software Intensive Systems, 2008. CISIS 2008. International Conference on 4-7 March 2008 Page(s):249 - 255