

Postprint: Applications and Research of Integrated Media Manuscript Technology

Authors: Song Lingyu, Mou Yunnan, Wang Wenkai

Date: 2023-10-08T00:00:00+00:00

Abstract

Composite media manuscript is a novel type of news information that supports block-based layout and semantic tag definition for multi-type news manuscript information content, such as text (leads, editor's notes), images (illustrations, accompanying images), audio-video, and other applications, thereby meeting the processing requirements of information resources in the present and future media industry.

Full Text

Composite Media Manuscript Technology: Applications and Research

Abstract: The composite media manuscript represents a new category of news information format that supports block-based layout and semantic tag definitions for diverse news content blocks—including text (leads, editor's notes), images (illustrations, inline images), audio-video, and other applications—thereby meeting the processing requirements of current and future media industry information resources.

Keywords: composite media manuscript; semantic tags; layout; publishing

Classification Number: G202

Document Code: A

Article ID: 1671-0134(2018)02-049-02

DOI: 10.19483/j.cnki.11-4653/n.2018.02.016

Authors: Song Lingyu, Mou Yunnan, Wang Wenkai

To address these challenges, we propose a new manuscript resource type—the “composite media manuscript.” This paper analyzes the characteristics of composite media manuscripts, identifies existing problems, and proposes corresponding solutions.

1. Current Status of News Resource Formats and Processing

Currently, news manuscripts in the media industry are primarily presented as text, images, audio-video, multimedia, and other application types. Relationships between manuscript resources are established through technical means such as grouping, special topics, packaging, and strong association, which are then leveraged for layout and multi-format publishing. However, as journalists' professional competence improves and new media formats emerge, the limitations of current manuscript processing methods in news production workflows—including gathering, editing, layout, and publishing—have become increasingly apparent, failing to meet user needs. These issues manifest primarily in the following aspects:

1.1 Independent Manuscript Types

Some media industry software processes different manuscript types—text, images, audio-video—independently. Related manuscripts are stored using layout/category classification or folders for subsequent layout and publishing. This creates forced constraints between software systems, requiring journalists to interact multiple times during news production to reach consensus, which may result in omissions or incomplete information in the final published content.

1.2 Difficulty in Maintaining Resource Relationships

For manuscript resources connected through grouping, packaging, special topics, and strong association, the workflow across editing, layout, and publishing software involves operations such as splitting/merging groups, splitting/merging packages, removing/joining topics, and dissolving/establishing associations. Any failed operation may cause confusion or loss of relationships, leading to cumbersome functional processing and difficult troubleshooting, thereby increasing maintenance overhead.

1.3 Complex Layout and Publishing

During layout, typesetters handling independently-typed manuscript resources must focus not only on page design but also on the relationships and sequencing between manuscripts. Simultaneously, they may need to understand the artistic conception that writers or editors wish to express while also handling templates for various manuscript types. This increases typesetter workload and delays publication time.

Although multimedia manuscripts can address multi-type information display, processing becomes complex when extracting content resources on demand for multi-purpose use (i.e., publishing to both traditional print and new media). These relationships are not easily maintained, and inconsistent tag definitions for

multimedia manuscripts across different software inevitably complicate system interoperability.

2. Design of Composite Media Manuscript Type

The composite media manuscript design enables text (leads, editor's notes), images (illustrations, inline images), audio-video, and other information types to be structurally displayed and associated within a single resource through distinct tag definitions, while also expressing the writer's intended layout and artistic conception. The design principles are as follows:

2.1 Semantic Label Design

2.1.1 Shoulder Title The shoulder title appears at the beginning of an article, before the main title. Format: begins with ## at the start of a line; no spaces allowed before ##; spaces immediately after ## are optional but meaningless and are not treated as content.

2.1.2 Main Title The main title appears after the shoulder title and before the subtitle. Format: begins with # at the start of a line; no spaces allowed before #; spaces immediately after # are optional but meaningless and are not treated as content.

2.1.3 Subtitle The subtitle appears after the main title and before the lead. Format: begins with ## at the start of a line; no spaces allowed before ##; spaces immediately after ## are optional but meaningless and are not treated as content.

2.1.4 Lead/Editor's Note The lead or editor's note appears immediately after the title. Format: begins with > at the start of a line; no spaces allowed before >; spaces immediately after > are optional but meaningless and are not treated as content.

2.1.5 Photo Group The photo group appears after the lead and before the main text. Format: begins with ![] (at the start of a line, followed by the unified file address of the image, then a closing parenthesis), after which comes the caption content. No spaces are allowed before !.

2.1.6 Inline Image in Main Text Inline images in the main text are denoted by ![] (beginning with the unified file address of the image and ending with a parenthesis).

2.2 Main Text Elements

2.2.1 Main Text Illustrations Main text illustrations appear in the main text, beginning with ![] (at the start of a line, followed by the unified file

address of the image, then a closing parenthesis), after which comes the caption content. No spaces are allowed before !.

2.2.2 Video Video appears after the photo group. Format: begins with ![video] (at the start of a line, followed by the unified file address of the video, then a closing parenthesis), after which comes the video description. No spaces are allowed before !.

2.2.3 Application Type Application type appears after the photo group. Format: begins with ![type] (at the start of a line, followed by the unified file address of the application type, then a closing parenthesis), after which comes the application type description. No spaces are allowed before !. The Type value can be any of the following: application x-world.

2.2.4 Audio Audio appears after the photo group. Format: begins with ![audio] (at the start of a line, followed by the unified file address of the audio, then a closing parenthesis), after which comes the audio description. No spaces are allowed before !.

2.2.5 Floating Title Floating titles appear in the main text, beginning with ### at the start of a line; no spaces allowed before ###; spaces after ### are optional and are treated as content. Two consecutive floating titles are processed as a single floating title, separated by
.

2.2.6 Emphasis Emphasis appears in the main text, denoted by content enclosed between ** and **.

2.3 Writing and Preview Design

We developed a semantic tag-supported editor for composing composite media manuscripts. This editor allows manuscript editing in plain text format, using special symbols to mark knowledge items such as titles, shoulder titles, subtitles, leads, and floating subheadings, thereby enabling knowledge-based tagging of all-media manuscripts for automatic extraction by backend systems and template-based display. Featuring a user-friendly interface, it uses a composite media galley that mixes text, images, audio-video, expressive titles, and captions as a unified content format. The composite media galley fully supports the UNICODE standard, providing support for Chinese, English, minor languages, and ethnic minority languages. It supports basic editing functions including copy, paste, delete, cut, select all, undo, and redo; common operations such as font and size modification; find and replace functions; real-time editing with immediate preview and transmission; preservation of revision marks and direct publishing channels; seamless integration with gathering and editing systems; and two editing modes: clean copy and draft. In draft mode, deletion and insertion marks from the current review round can be visually inspected, with

different background or font colors used to distinguish current modifications, meeting the writing needs of composite manuscripts for the present and near future. Composite media manuscript content can be previewed in the editor or transformed into HTML format for third-party system access.

2.4 Layout and Publishing Design

Layout Support: (1) Direct input into layout software that supports composite media manuscripts, where the software parses semantic definitions and automatically typesets content on demand using template-based techniques, enabling “instant layout.” (2) For layout software that does not support composite media manuscript semantics, the manuscript must be converted into standard XML, HTML, and other formats for the layout software to process according to its own capabilities.

Publishing Support: (1) Composite media manuscripts can be converted into HTML and XML outputs required by downstream publishing systems, adapting to their templates. (2) Composite media manuscripts support previewing of publishing results.

This paper identifies the importance of designing a new manuscript type by examining the limitations of current news manuscript types and emerging needs in the media industry. It proposes principles and specifications for the semantic definition, writing, editing, preview, layout, and publishing of composite media manuscripts, thereby designing a new manuscript type standard that meets new user requirements, is suitable for displaying and associating multi-type news information, and works for both traditional and new media, in order to improve practitioners’ work efficiency and news production quality.

Funding: This research is supported by the National Digital Composite Publishing System Engineering—Media Collaborative Work System (Project No.: 1141STC40519/07).

References: [1] Yin Shengzhong. Technical Construction Scheme for Multimedia Gathering and Editing System Under Media Convergence Development [J]. China Media Technology, 2016, 7.

[2] Cheng Yitao. The Transformation Path of News Gathering and Editing in the New Media Environment [J]. News Research Guide, 2015, 20.

[3] Wang Rong. Analysis of Using HTML Online Text Editor in Java Web Development—Taking KindEditor as an Example [J]. Journal of Hubei University of Science and Technology, 2013, 33.

(Author affiliations: Weifang Beida Jade Bird Huaguang Phototypesetting Co., Ltd.)

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.