A Structured Paper Writing Tool for Semantic Publishing: Design and Implementation Post-print Abstract: This paper proposes a structured paper writing tool for semantic publishing, aiming to address the problems of lacking semantic structure and poor machine readability in traditional academic wri···

Authors: Le Xiaoqiu, Wang Zixuan, Zhang Xiaolin, He Yuanbiao, Fu Changlei, Xu Liyuan

Date: 2017-11-08T00:00:00+00:00

Abstract

【Objective】 To construct a paper writing tool for semantic publishing, enabling content structuring and objectification during the writing phase, such that a paper itself becomes a system that is runnable, interactive, and experiential.

【Methods】 Digital object and digital template technologies are employed to decompose paper content (metadata, sections, data, rich media, etc.) into different types of digital objects, which are organized through templates. Interaction is achieved via an event-triggering mechanism, and editing and presentation are performed in HTML5 web page format, with storage as an XML structured document package.

【Results】The DPaper structured paper writing tool has been launched, providing a series of functions ranging from material collection (cloud notes), digital object creation, automatic reference indexing, journal layout presentation, to Word document format conversion. Paper content has achieved objectification and partial semanticization.

【Limitations】 Compared with conventional paper editors, the digital object editor' s functionality is still incomplete, unable to create objects such as formulas and graphics, and lacks sufficient typesetting flexibility.

【Conclusion】 Using the DPaper writing tool, authors can construct structured papers that meet the application requirements of semantic publishing during

the writing stage.

Full Text

DPaper: Design and Implementation of a Structured Paper Authoring Tool for Semantic Publishing

Le Xiaoqiu[1], Wang Zixuan[1,2], Zhang Xiaolin[1], He Yuanbiao[1], Fu Changlei[1], Xu Liyuan[1]

[1](National Science Library, Chinese Academy of Sciences, Beijing 100190, China)
[2](University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract

[**Objective**] This study develops a paper authoring tool for semantic publishing that enables content structuring and objectification during the writing phase, transforming each paper into a system that is executable, interactive, and experiential. [**Methods**] We decomposed paper content (metadata, chapters, data, rich media, etc.) into different types of digital objects using digital object and digital template technologies. These objects were organized through templates and interacted via event trigger mechanisms. The system employed HTML5 web pages for editing and presentation, storing documents as XML-structured document packages. [**Results**] The DPaper structured paper authoring tool is now online at http://idpaper.las.ac.cn, providing a suite of functions from material collection (cloud notes) and digital object creation to automatic reference indexing, journal layout rendering, and Word document conversion. Paper content has achieved objectification and partial semanticization. [**Limitations**] Compared with conventional paper editors, the digital object editor' s functionality remains incomplete, lacking capabilities to create formulas and graphics, and offering insufficient layout flexibility. [**Conclusions**] Using the DPaper authoring tool, authors can construct structured papers that meet semantic publishing application requirements during the writing stage.

Keywords: DPaper; Semantic Publishing; Structured Paper; Digital Object; Authoring Tool

---

1. Introduction

Digital environments have introduced significant new trends in the morphology and utilization of academic papers, including content structuring, objectification, and semanticization. These digital-native paper formats enable computability and enrichment (incorporating interactive data viewers, graphics/images, charts, visualizations, and "living equations"), potentially transforming publishing into

a software model that provides better representation for both authors and researchers while enabling continuous innovation in fine-grained content applications. Semantic publishing, as an emerging application paradigm, has entered a rapid development phase in recent years, facilitating easier data integration between papers, enabling identification and indexing of knowledge objects and relationships, and incorporating parsing logic and results as organic components of content publishing. STM' s 2015 technology trends report posits that journal articles occupy the central position in a "Hub and Spoke" publishing model, connecting videos, graphics, tables, and various digital artifacts, with data ascending to become a primary research object. However, no truly usable paper authoring tool currently exists to assist authors in generating structured papers that satisfy semantic publishing requirements during the creation phase.

To address this gap, we developed DPaper (http://idpaper.las.ac.cn/), a structured paper authoring tool oriented toward semantic publishing. DPaper fundamentally transforms paper utilization patterns by achieving content structuring and objectification during the writing stage, turning papers into software models where each paper becomes a system that is executable, interactive, and experiential. This approach enables readers to manipulate and reuse authors' research data, processes, and results. This paper elaborates on the tool' s primary research concepts and system design implementation methods.

2. Related Work

Recent explorations in structured paper research for semantic publishing have primarily focused on content objectification and semanticization through three main approaches: paper content modularization, digital object encapsulation with semantic description, and semantic annotation. Representative models include the modular paper model and the semantic publishing model. While semantic annotation constitutes the primary means of literature content semanticization with numerous studies and experiments, this paper does not focus on that aspect.

The modular paper model, proposed by Kircz, conceptualizes papers as composed of modules defined as information units with unique characteristics and self-contained conceptual representations. Datasets, images, audio, video, and other elements are treated as independent yet interactive objects or modules aggregated into papers, connected into fixed units for communication purposes. This modular structure enhances reading and publishing efficiency, as demonstrated in Cell' s paper structure. Applying digital objects to organize dissertations represents a typical application of modular thinking, integrating digital object applications into existing electronic thesis and dissertation systems while providing METS/XML conversion and import/export functions. The OpenETD tool exemplifies this approach, serving as both a standalone dissertation submission system and a component for institutional repositories using METS/XML export functionality. In the ProQuest/UMI system, rich media such as audio, video, and datasets (spreadsheets) are submitted online as sup-

plementary files with corresponding descriptive information.

In semantic publishing models, Hunter proposed the Scientific Publication Package (SPP), a new information format for encapsulating raw data, derived products, algorithms, software, text, relevant context, and metadata. SPP enables scientists to acquire, index, store, share, exchange, reuse, compare, and integrate scientific results. Based on numerous scientific concept models, SPP is a composite digital object represented as an RDF package, with relationships between internal atomic objects either explicitly defined through ontology rule reasoning during metadata acquisition or defined by scientists during SPP description. The model emphasizes workflow technology as a component of scientific processes for capturing processing step chains that generate scientific data and derived products, enabling scientists to describe and execute their experimental processes in a reproducible, verifiable, and distributed manner while tracking error sources and processing defects.

Regarding structured paper editing tools for semantic publishing, no general-purpose tools currently exist. However, the BioLit project and the SCOPE (Scientific Compound Object Publishing and Editing System) project have conducted valuable explorations from semantic markup and composite digital object perspectives. In the BioLit project, Fink et al. developed an XML-based writing tool using the National Library of Medicine' s Document Type Definition (NLM DTD) to store standardized, machine-readable publications. This DTD includes semantic markup and unique identifiers for articles and object content (such as figures and tables), facilitating integration of open literature and biological data, tested using the entire PLoS and Protein Data Bank (PDB) corpora.

The SCOPE tool represents an attempt to enable researchers to construct digital objects themselves, which is the ultimate solution for structured papers in semantic publishing, as only authors fully understand their specific research processes, computational methods, experimental materials, data, and results. SCOPE is a digital content concatenation tool utilizing the OAI-ORE specification, designed as a scientific compound object publishing and editing system to enable scientists to easily create, publish, and edit scientific compound objects, encapsulating different datasets and resources from scientific experiments or discovery processes for individual compound object publication and exchange. However, SCOPE construction examples demonstrate that building composite objects requires semantic web relationships, making it difficult even for ICT experts to complete, with usability and practicality challenges that limit current adoption.

3.1 Conceptual Model for Semantic Publishing Authoring Tools

Current paper writing predominantly employs document editors (such as Word), where content exists as static composite documents (e.g., Doc/PDF formats) that are unstructured, non-semantic, and weakly interactive. This makes it difficult for authors to completely and effectively present their research data

and processes to readers, while research results remain challenging for peers to effectively utilize, understand, observe, and verify.

A semantic publishing-oriented paper authoring tool represents a computable, reusable/verifiable, and interactive paper system with operable, composable, and publishable content featuring multiple presentation modes. Papers created with such tools should possess four key capabilities: (1) executability, where a paper can be published as an application system; (2) digital object representation with structured and semantic content; (3) rich media objects that fully demonstrate scientific research processes and achievements; and (4) independently executable digital objects that can be combined and customized to meet semantic publishing requirements. DPaper incorporates the modular paper model concept in its design.

4.1 Organization of Paper Digital Objects

Paper content is decomposed and described by granularity, with standardized descriptions referencing METS, Dublin Core, and the Book and Collection Tag Library version 3.0 developed by NCBI and NLM [Figure 2: see original paper]. By constructing standardized digital paper templates, dissertation content organization is separated from presentation, transforming academic papers from static composite documents into configurable, operable, transferable, exchangeable, and preservable digital object collections. Digital objects are interrelated, with content organization described using open metadata standards (such as METS and Dublin Core) and presentation displayed through web formats. For digital object operations, corresponding processing specifications and interface standards are established, converting some composite digital objects into integrable microservices.

3.2 System Processing Framework

The DPaper system framework consists of five components [Figure 1: see original paper]: (1) digital document representation, which uses description specifications to formulate corresponding digital paper templates; (2) digital object production, responsible for object data management, object interaction, encapsulation, and data conversion; (3) digital document editing, the venue for digital paper creation, editing, and modification, organizing digital objects based on paper structure units and assigning semantic tags to objects during editing to achieve structuring and semanticization; (4) personalized composition of digital objects in the editor according to templates, with inter-object data associations presented as web pages for browsing and publishing; and (5) storage of digital objects and their data as web packages, with structural description information stored in XML data files for digital document exchange and third-party software reuse.

4.2 Communication and Interaction Between Digital Objects

The DPaper system integrates digital object set data loading, processing, editing, presentation, and storage operations, enabling mutual data invocation between digital objects where data operations in one object can immediately trigger state and result updates in another object during editing or execution. This interaction primarily occurs through event trigger mechanisms, where the system notifies corresponding digital objects to update their responses when users perform certain operations.

DPaper defines multiple response events for digital objects, including create, begin modification, end modification, delete, copy, and erase. The event processing flow follows: document operation (button/shortcut) → environment processing → trigger Before event for permission checking and related operations → trigger Manager response event → trigger After event → complete event. The data object basic event (create, modify, delete) response process is illustrated in [Figure 3: see original paper].

4.3 Semantic Description of Digital Objects

Digital template technology marks corresponding semantic tags for different granularities of digital objects defined in digital papers, automatically generated by the system. To reduce editing complexity, the system currently does not perform more detailed semantic processing of digital object content. In DPaper, we constructed two types of digital templates using the "Chinese Academy of Sciences Master' s Thesis" template and the *Modern Library and Information Technology* journal paper template. Templates use elements defined in Section 4.1 to mark digital objects of different granularities (such as cover, statement, title, author, institution, chapters, figures, tables, references, etc.), with specific definitions for font, size, style, position, composition, and format recorded as XML files in the system for invocation during paper object editing.

[Figure 4: see original paper] and [Figure 5: see original paper] present semantic description examples of dissertation cover objects and reference objects, respectively.

4.4 Reuse Mechanisms for Dissertation Digital Objects

DPaper currently provides three reuse modes: data reuse, digital object reuse, and whole paper reuse.

(1) Data Reuse: Data within digital objects achieves reuse through format conversion. Data in digital objects such as Dtable, Dchart, and relationship diagrams can be converted into JSON, CSV, RDF/XML, and other format data files.

(2) Digital Object Reuse: Data, programs, and library files within digital objects are encapsulated into independent web packages with access entry points and object description metadata, capable of running independently in browsers

outside the DPaper environment. After copying, objects can be embedded into other digital papers for reuse.

(3) Whole Paper Reuse: Within the system, DPaper uses three XML files: a paper structure file recording hierarchical relationships between paper objects, a data file recording paper metadata and digital object content, paths, and locations, and a display format file recording display information for various digital objects in standardized display templates, such as font, size, position, color, and style. These internal structure files are converted into discrete METS electronic documents for digital preservation or document exchange, as shown in [Figure 6: see original paper].

5.1 DPaper Software Composition and Main Functions

DPaper comprises three components: a paper editor, cloud notes, and a Word plugin [Figure 7: see original paper]. The paper editor is desktop software serving as the primary platform for structured paper construction, responsible for paper creation, digital object creation, object content editing, data management, object management, web preview, and document conversion. Cloud notes facilitate material collection and collaborative team writing, enabling note extraction across PC, mobile, and web platforms with data synchronization. The Word plugin constructs DPaper structured documents within the Word environment, enabling conversion between the two document formats.

5.2 DPaper Paper Construction Process

The DPaper structured paper construction process is illustrated in Figure 8: see original paper, while Figure 8: see original paper shows a screenshot of a Chinese Academy of Sciences master' s thesis generated using DPaper.

DPaper explores construction methods for structured paper authoring tools oriented toward semantic publishing, designing and implementing a practical software system for the entire paper writing lifecycle. Using a digital template mechanism, papers are represented as three interrelated yet separate structured files: data, structure, and presentation style, achieving content structuring and partial semanticization during the writing phase. The introduction of rich media objects enhances paper operability and reusability, enabling the generation of Word documents, machine-readable XML documents, or executable web document systems according to specific applications. This approach realizes paper structuring and partial semanticization at the creation source, which is significant for advancing semantic publishing.

DPaper still has limitations: its main platform' s conventional editing functions lag considerably behind Word, cannot create objects such as formulas and graphics, and requires practical validation of whether authors accustomed to Word and WPS will adapt to this new editing paradigm. Additionally, digital objects cannot be smoothly invoked between the main platform and Word plugin, requiring stability improvements. Future work will address these issues through

corresponding system enhancements.

References

[1] Shotton D. Semantic Publishing: The Coming Revolution in Scientific Journal Publishing [J]. Learned Publishing, 2009, 22(2): 85-94.
[2] Zhang Xiaolin. The Forces Disrupting Digital Library [J]. Journal of the Library Science in China, 2011, 37(5): 4-12.
[3] From STM, Tech Trends for 2015 [EB/OL]. [2016-10-11]. http://beyondthebookcast.com/from-stm-tech-trends-for-2015/.
[4] STM Tech Trends 2014 [EB/OL]. [2016-10-11]. http://www.stm-assoc.org/2014_{{{04}}{{29}}}{{{Innovations}}{{USA}}}{{{STM}}{{Tech}}}{{{Trends}}{{2014}}}.pdf.
[5] Kircz J G. Modularity: The Next Form of Scientific Information Presentation? [J]. Journal of Documentation, 1998, 54(2): 210-235.
[6] Kircz J G. New Practices for Electronic Publishing 2: New Forms of the Scientific Paper [J]. Learned Publishing, 2002, 15(1): 27-32.
[7] OpenETD: Open Source Electronic Theses and Dissertations Management Software [DB/OL]. [2015-08-06]. https://rucore.libraries.rutgers.edu/open/projects/openetd/index.php.
[8] ProQuest Dissertation Publishing [DB/OL]. [2015-08-06]. http://www.etdadmin.com/UMI{PreparingYourM
[9] Hunter J. Scientific Publication Packages-A Selective Approach to the Communication and Archival of Scientific Output [J]. Journal of Digital Curation, 2006, 1(1): 3-16.
[10] Enhanced Publications [EB/OL]. [2016-10-11]. http://www.doc88.com/p-873117284280.html.
[11] Fink J L, Bourne P E. Reinventing Scholarly Communication for the Electronic Age [J]. CTWatch Quarterly, 2007, 3(3): 26-31.
[12] Cheung K, Hunter J, Lashtabeg A, et al. SCOPE: A Scientific Compound Object Publishing and Editing System [J]. International Journal of Digital Curation, 2008, 3(2): 4-18.
[13] Book and Collection Tag Library Version 3.0 [EB/OL]. [2016-10-11]. http://dtd.nlm.nih.gov/book/tag-library/3.0/index.html.

Author Contributions

Le Xiaoqiu: Proposed the research concept for structured papers in semantic publishing, designed the research methodology, oversaw system implementation, and wrote and revised the paper.
Wang Zixuan: Conducted literature review, software development and testing, and paper revision.
Zhang Xiaolin: Proposed the Smart Dissertation research direction and goals, and revised the paper.
He Yuanbiao: Responsible for system development and solving key technical problems.
Fu Changlei, Xu Liyuan: Conducted module development and testing.

Conflict of Interest Statement

All authors declare no conflict of interest.

Received: September 13, 2016
Revised: October 19, 2016

DPaper: A Structured Paper Authoring Tool for Semantic Publishing

Le Xiaoqiu[1], Wang Zixuan[1,2], Zhang Xiaolin[1], He Yuanbiao[1], Fu Changlei[1], Xu Liyuan[1]

[1](National Science Library, Chinese Academy of Sciences, Beijing 100190, China)
[2](University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: [Objective] We developed a paper authoring tool for semantic publishing, which makes the article' s content structured and object-oriented. Each paper is a system with executable, interactive and experiential features. [Methods] First, we divided the content of each paper (metadata, chapters, data, media etc.) into objects organized by digital template. Second, these elements interacted with each other through the event trigger mechanism. Finally, the paper was modified and presented with HTML5 pages, and then, saved as XML documents. [Results] DPaper is available at iDPaper.las.ac.cn, which provides a series of functions such as material collection (cloud notes), digital object creation, automatic reference indexing, Word document format conversion in accordance with periodical layouts etc. The paper' s content is object oriented and partial semantization. [Limitations] Compared to conventional paper editors, the DPaper' s digital object editor could not create formulas or graphics, and is not flexible to change layouts. [Conclusions] DPaper could help us compose a structured paper that meets the requirements of semantic publishing.

Keywords: DPaper; Semantic Publishing; Structured Paper; Digital Object; Authoring Tool

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