An Ontology-Based Platform for Scientific Writing and Publishing

Hao Xu[1,2,3], Changhai Zhang[1], Yang Ai[1,3], Ziwen Wang[1], and Zhanshan Li [1,*]

[1] College of Computer Science and Technology, Jilin University,
Qianjin Street 2699, Changchun 130012, China
[2] Center for Computer Fundamental Education, Jilin University,
Qianjin Street 2699, Changchun 130012, China
[3] Department of Information Engineering and Computer Science, University of Trento,
Via Sommarive 14, Povo 38123, Italy
hao@disi.unitn.it, changhai@jlu.edu.cn,
aiyangjilin@yahoo.com.cn, mierkelin@sina.com,
zsli@email.jlu.edu.cn

Abstract. Writing scientific discourses and publishing academic results are integral parts of a researcher's daily professional life. Although tremendous magic have been brought by advancement of digital library technologies and social networking services, there are still no off-the-shelf utilities for strategic writing, reading and even publishing. In this paper, we propose an ontology-based platform for scientific writing and publishing featured with its semi-automatic metadata generation and semantic-linked composition, aiming to facilitate the efficient creation, dissemination and reuse of scientific knowledge.

1 Introduction

Since the Elsevier's Journal *Cell* launched the new online publishing format "Article of the Future" [1] in the year of 2010, a revolution in scientific publishing has taken place again. In recent years, a handful of models were proposed for scientific discourse representation grounded on the Rhetorical Structure Theory (RST) [1] such as ABCDE Format [2], SALT (Semantically Annotated LaTex) [3], etc. Also, several applications of ontology-aware text mining and retrieval systems, e.g. Textpresso [2] and iHOP [3] are widely accepted in particular domains of biological science.

Ontology-based structure for representing scientific papers, along with semantic web techniques for linking related entities and concepts, forms the basis for new types of scientific writing and publishing. Within the construction of our proposed ontology, we also consider defining the patterns of papers' logical structuring for strategic writing, reading, and the interoperability with other domain ontologies as well as metadata schemes. The advantage of such proposition also evolves the

* Corresponding author.
[1] "Article of the Future": http://beta.cell.com/
[2] Textpresso: http://www.textpresso.org/
[3] iHOP: http://www.ihop-net.org

T.-h. Kim et al. (Eds.): FGIT 2010, LNCS 6485, pp. 267–271, 2010.

collaborative operations in the research communities including metadata annotation, maintenance, sharing and so forth.

In this paper, we briefly introduce our ongoing project [4, 5, 6, 7] of an ontology-based platform for scientific writing and publishing. The objective of this project aims at applying rhetorical structure theory to online authoring, navigating and annotating. Thus readers can access to their own interests directly without being overwhelmed by the extra noisy information. The previous indivisible linear-structured papers are re-organized by ontology-based structure within meaningful semantics.

2 Overview

This platform consists of two main components. One is an *Editor* used for semantically authoring, where "semantically" means that the system may offer the functionality of encoding various semantic links. For instance, in traditional system one is difficult to track versioning during the manuscript's evolving. Semantic links do definitely solve these problems via sets of URLs and metadata information. Moreover, various reviews and comments associated to different draft versions are also considered when the Editor is designed. We call it lifecycle management which is one of functions in the Editor. Others services like local metadata, writing patterns, strategic reading, ubiquitous reuse and semantic search are supported.

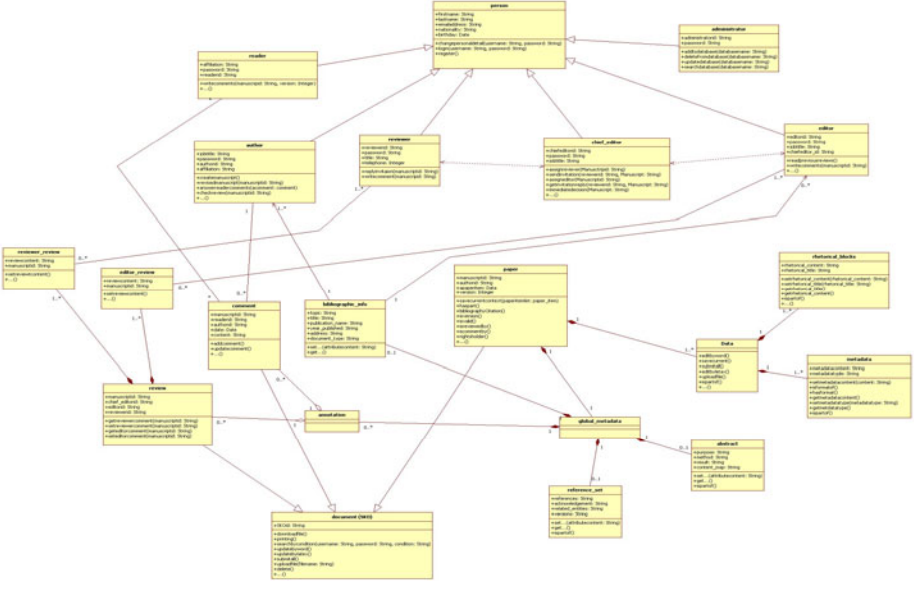

Fig. 1. UML Class Diagram for Ontology Development

The other one is an *Online Portal* for navigating. More rhetorical structure templates go beyond the domain-specific prototypes of existing online publishing systems. Types of document and roles of person are assigned specifying metadata

schemes for faceted search. Social networking makes the community collaborative and communication efficient.

Herein, an overview of platform design is illustrated in Figure 1. This UML class diagram shows the core of ontology development, which is as well as the core of whole platform. Document (Paper, Review, and Comment) and Person (Author, Reader, Editor, Chief Editor, and Administrator) constitute the main classes, along with sets of attributes and operations. Sub-patterns are implemented for each types of document, i.e. paper, review and comment, likewise. This methodology aims to divide linear papers into parts, namely rhetorical chunks. Instead of ordered sections and subsections, there rhetorical chunks are connected via types of semantic links mentioned above. One paper is composited by Global Metadata and Data, while Data is organized by several rhetorical blocks and multiple local metadata related with each rhetorical block. All metadata can be accessed independently, which we believe will make metadata more useful and accessible to readers.

More detailed functional design and implementation for Editor and Portal as ongoing work will be specified and deployed in the near future.

3 Discussion

The *"Article of the Future"* breaks through the traditional structure of papers. While the papers of Ontology-based platform are based on the structure features of *Cell,* they employ more semantic web techniques to each rhetorical chunk.

Table 1. Structural Comparison with Article of the Future and Traditional Article

Article / Features	**Traditional Article**	**Article of the Future**	**Article of Ontology-based Platform**
Paper structure	indivisible linear structure	divisible linear structure	divisible Ontology-based rhetorical structure

As illustrated in Table 1, we can tell the revolution in paper structure's evolution. In an ontology-based structured article, A*uthor* can provide metadata for every rhetorical chunk, and then an *Editor* may adopt the design methodology of supplemental semantic information to data and other materials that support or relate to main conclusions of each rhetorical chunk. But this supplemental information is considered as additional information or references. So divisible ontology-based structure can help readers explore papers conveniently and abundantly.

On semantic web, ontology plays a most important role. Because it strictly defines concepts and relations of such concepts, it can promote knowledge sharing and reusing in the knowledge level. Table 2 shows the article of ontology-based platform with features of introducing multiple types of data and semantic links to support the conclusions of the paper. In these papers, abundant data information (figures, audio resources, movie resources, descriptive text, etc) need to be constructed in the ontology to apply metadata describing these resources. What's more, the abundant

ontology-based semantic search can generate semantic links and concepts, which provides multiple metadata to both whole paper and paper parts.

Table 2. Data Feature Comparison with Article of the Future and Traditional Article

Article / Data / Feature	**Traditional Article**	**Article of the Future**	**Article of Ontology-based Platform**
Figures and tables	main text (e.g. by pdf file)	main text & Individual supplemental information(by multiple formats file)	main text& Individual supplemental information (by multiple formats file, Semantic links, etc)
Experimental procedures	main text (e.g. by pdf file)	main text & Individual supplemental information (by pdf file, multimedia files)	main text & Individual supplemental information (by xml file, multimedia files)
References	Based on whole paper (e.g. by pdf)	Based on whole paper& each section (by pdf)	Based on whole paper& each section (by xml, Semantic entities and concepts links)
Readers' comments	None	Based on whole paper	Based on whole paper or each section
Multimedia files	None	Movies, Audio Clips, etc	Movies, Audio Clips, Semantic links, etc

Acknowledgement

This research is partially supported by China Scholarship Council (CSC). And many thanks to Prof. Fausto Giunchiglia for his inspiring suggestions.

References

[1] Thompson, S.A., Mann, W.C.: Rhetorical structure theory: A theory of text organization. Technical report, Information Science Institute (1987)
[2] de Waard, A., Tel, G.: The abcde format enabling semantic conference proceedings. In: SemWiki (2006)
[3] Groza, T., Handschuh, S., Mönller, K., Decker, S.: Salt – semantically annotated latex for scientific publications. In: Franconi, E., Kifer, M., May, W. (eds.) ESWC 2007. LNCS, vol. 4519, pp. 518–532. Springer, Heidelberg (2007)
[4] Xu, H., Zhang, C.: Towards a pattern-based approach for scientific writing and publishing in Chinese. In: Chowdhury, G., Koo, C., Hunter, J. (eds.) ICADL 2010. LNCS, vol. 6102, pp. 264–265. Springer, Heidelberg (2010)

[5] Xu, H.: A semantic pattern approach to managing scientific publications. In: Kim, T.-h., Adeli, H. (eds.) AST/UCMA/ISA/ACN. LNCS, vol. 6059, pp. 431–434. Springer, Heidelberg (2010)
[6] Xu, H.: Managing ubiquitous scientific knowledge on semantic web. In: Kim, T.-h., Adeli, H. (eds.) AST/UCMA/ISA/ACN. LNCS, vol. 6059, pp. 421–430. Springer, Heidelberg (2010)
[7] Xu, H.: A pattern-based representation approach for online discourses. In: Kim, T.-h., Adeli, H. (eds.) AST/UCMA/ISA/ACN. LNCS, vol. 6059, pp. 378–384. Springer, Heidelberg (2010)