An Analysis of Platforms for Scholarly Publication

Yun Li
41017919
Department of Computing
Macquarie University
Sydney, Australia
yun.li1@students.mq.edu.au

Abstract

In the journal publishing world, there involves various text processing tasks such as copy-editing or revision. Modern text processing technologies, and in particular the kinds of techniques developed in natural language processing and language technology, make it possible to automate some of these editing and revision tasks that are currently performed by authors and editors. However, it is difficult for these technologies to fit well with the variety of different text processing platforms and wolkflows of publishing process. The aim of this paper is to help us understand the range of ways in which authors and publishers are currently work and tools that they use, then to identify what kinds of integration would be likely to be most successful. We summarise a general editorial workflow from a web observation. We also conduct a survey to gather information of production process in which relevant data is hard to find. Base on the 98 responses from the survey, we outline 3 most common production workflows. For each workflow, as well as the editorial workflow, we finally indicate the target audience and some possible automated editing assistance which might be introduced to them.

1 Introduction

Scholarly journal publishing is usually described as distribution of academic research and scholarship. Most academic work is published in journal article, book or research paper form. The process of journal publication, especially in scholarly publication, involves various text processing tasks such as peer-review, copy-editing, typesetting and proofing. Although some of these tasks have been streamlined by some modern technologies, such as manuscript management systems for managing peer-review workflow, a number of editing and revision tasks are still labor-intensive as they are performed by authors, editors and publishers.

In 2003, Mabe (2003) indicated that there were approximately 1.4 million articles published by about 21000 journals. As the number of researchers is increasing, Steele (2006) showed the amount of journals has accrued to 23000. On behalf of these journals, it will be genuinely useful if some modern text processing technologies could be applied to automate some of the labor-intensive tasks in journal publication; Dale (1997) also indicates that speech and natural language processing technologies are likely to be an Information Technology solution which could assist in these tasks. However, the variety of different text processing platforms used by authors, editors and publishers and different workflows of publishing process make it difficult to determine what kinds of integration would be likely to be most successful. We are particularly interested in the workflows of production process since the general editorial process is already known. So the purpose of this paper is to carry out an analysis of the general workflows of publishing process and the tools that are being used in these workflows, then identify places where automated editorial assistance might be introduced and recommend some kinds of possible automations. The approach we use for information gathering is a survey of journal publishing process. We select 400 journal editors[1], which mostly are editor-in-chief and production editor, from 360 A+, A and B ranked computer science journals on the CORE journal list[2]. The survey does not consider the journal editorial process as the information of this process can be acquired by the web observa-

1 Some journals have cooperate editors-in-chief
2 http://www.core.edu.au/

tion. Instead, the survey has been undertaken to better understand the production process in which the information is hard to find.

In section 2, we intend to review some general journal publishing process in which authors and editors currently work, and the tools and text processing platforms that are normally used. In section 3, we introduce the survey of journal publishing process along with its results. Then an analysis of these results is carried out in section 4. In section 5, we suggest some possible integrations of automated editorial assistance that might be introduced in scholarly journal publishing process.

2 General Publishing Process

Recently, scholarly publishing is undergoing major changes as it is transferring from paper-based to electronic format. The paper-based publishing has persisted for a long time as it is easy for authors to produce submissions; for editors and reviewers, there is no special technology required to view submissions. During these years, since the increasing costs in handling paper submissions and the largely use of Internet, paper-based publishing starts to go electronic. However, the main processes involved in traditional scholarly publishing and electronic scholarly publishing are almost the same. As mentioned by Campbell *et al.* (1997), the key processes can be described as editing, production, marketing, distribution, sale and promotion. This paper focuses on the two earliest stages: editing and production, since they involve most of the text processing procedures.

2.1 Journal Editing Process

Editorial process, which is mainly described as the process of peer-review in scholarly journal publishing, covers the following common stages as indicated by Ciesielski (2005):

- Editors collect submissions from authors
- Peer-review among editors, reviewers, and authors
- Editors make decisions (reject, accept, correct)
- Authors correct and submit

The process of peer-review is usually organized by the journal editor, then conducted by editors and reviewers, and completed when the article is accepted for publishing. A general peer-review process can be depicted in figure 1:

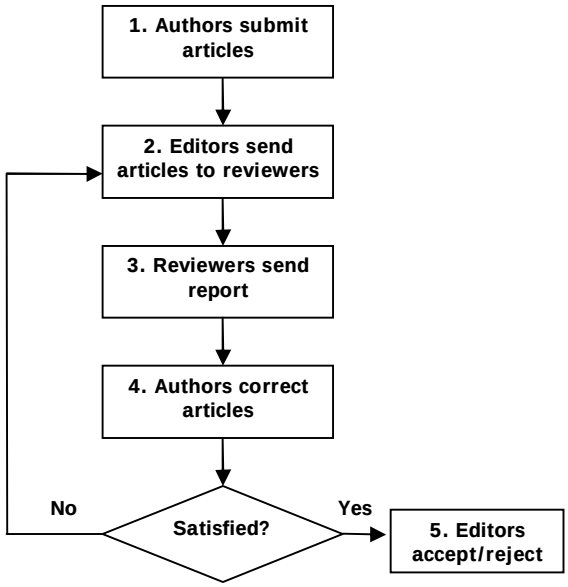

Figure 1 General Editorial Workflow

In order to better understand the peer-review process, we perform a web observation with journals on the CORE list. Approximately 100 A+ ranked journals' websites, which contain direct information about journal editing process, are selected for observation. The summary of results is quite similar to the workflow in figure 1: first of all, an author submits an article to a journal editor, then the editor checks whether the article is of sufficient quality to go through the peer-review. If the article is appropriate to the journal scope, then the editor assigns 3-5 reviewers who have the similar academic background as the author to each article for reviewing. After reviews are finished, reviewers send their reports on the article along with their comments back to the editor. Once the editor gets the reports, he/she asks the author to correct the article base on the comments and then in most situations, the article is asked to be reviewed again and again until the article is satisfied to the editor. After the peer-review process is done, the editor finally decides to accept or reject the article for publishing.

The electronic peer-review process can be automated through the use of online manuscript management systems such as ScholarOne ManuscriptCentral[3], Aries Editorial Manager[4], and EJournalPress[5]. These systems have been used to manage the submission and peer-review process effectively: authors can submit their

3 http://scholarone.com/products/manuscript
4 http://www.editorialmanager.com/homepage/home.htm
5 http://www.ejpress.com/index.shtml

manuscripts to the system; editors can review the submissions and assign reviewers directly in the system. The manuscript management systems assist editors to manage peer-review rapidly and comfortably; they automate or eliminate many manual steps in editorial process.

2.2 Journal Production Process

Once the submissions are finally accepted by the editors after the editorial process, the production process happens and it is controlled by a production editor or publisher. The production process covers the following common stages: copy-editing, typesetting, inclusion in a specific issue of a journal, and then printing and publication.

Copy-editing is performed by copy-editors to check the spelling and grammar errors of the article and make sure the referencing is correct. Typesetting is all about the presentation of an article: typesetters deal with layouts, fonts, headings etc to make sure the article is formatted in a manner that is consistent with the style of the journal. Proof-reading is involved for the author to review and correct proofs in one or more stages of the production process. As Dale (1990) indicated, because both copy-editing and proof-reading are time-consuming and error-prone processes, automated assistance would be most likely to happen there. However, information regarding the details of the production process is inadequate and hard to find through the web observation. We need to understand the production process well enough before deploying any automated editorial assistance.

2.3 Existing Editing Assistance Tools

Almost each one of the publishing processes we mention above involves the using of text processing tools: authors write their articles using Microsoft Word or LaTeX base on the demand of the journal they submit to; copy-editors enter corrections on the submissions using Microsoft Word or manually base on the formats of these submissions; typesetters format the submissions using Plain TeX, LaTeX or other tools. Microsoft Word and LaTeX both are standard word processing software. Word has a built-in spell checker and is easy to use whereas LaTeX provides a professional layout and faster typing speed than Microsoft Word (users only need to concern the content rather than the format in LaTeX).

Because the information about the production process is hard to find through the web observation and moreover, there are variety of platforms exist, it is difficult for us to determine what kinds of tools would be most beneficial, and what forms of integration would be likely to be most successful.

In order to better understand the production process and existing editing tools which have been utilized in scholarly publishing, we conducted a survey of journal publishing process which has been distributed to around 360 computer science scholarly journals and for each journal, we select one or two editors, including editor-in-chief and production editor. In the following section, we will discuss the survey and its results in detail.

3 Survey of Journal Publishing Process

The survey of journal publishing process acquired information from journal editors in the field of computer science. A 10-question web survey was e-mailed to approximately 400 editors from 30th Mar 2009 to 8th May 2009. These editors, including editors-in-chief and production editors, are selected from 360 A+, A and B ranked computer science journals on the CORE journal list. Of these 400 editors contacted, 98 responses (or a response rate of 24.5%) are aggregated for the analysis of this study.

The object of this survey is to help us understand the journal publishing process, especially the production stage. These 10 questions in the survey are designed to find out possible text processing platform(s) and workflow(s) that might exist in the production process.

The first question is a very general question which asks each participant to provide the name of his/her journal. 95 of the editors provided the information while only 3 of them skipped the question. These results are useful when we analyze on a journal by journal basis.

Question 2 and 3 acquire information about how accepted manuscript is processed. Question 2 asks about the role that is responsible for copy-editing in the production process. The results from this question are really important for us to find out our target audience of possible automation in copy-editing. As showed in table 1, approximately 47% of the journals have in-house

copy-editors who are responsible for the copy-editing work of accepted manuscripts. For 20% of the journals[6], copy-editing is done by editors or out-sourced.

Similar as question 2, the third one aims to find out the role that incorporates copy-editor's revisions into manuscripts for return to authors for checking. Table 2 indicates that 39.5% of copy-editor's revisions are done by copy-editors themselves, and around 20% are done by typesetters (both in-house and external). A small percentage (≤8%) of revisions is incorporated by authors.

Q2: How is copy-editing of accepted manuscripts carried out for your journal?	
copy-editors on the fulltime staff	46.9%
external freelance copy-editors	14.8%
no copy-editing is carried out	11.1%
other	27.2%

Table 1 Results of Question 2

Q3: How are the copy-editor's revisions incorporated into the manuscript for return to the author for checking?	
copy-editors	39.5%
in-house typesetters	11.8%
external typesetters	7.9%
author	7.9%
other	32.9%

Table 2 Results of Question 3

Q4: In which format(s) do you provide the accepted versions of manuscripts to your copy-editors?	
hard copy	6.9%
electronic-word	62.56%
electronic-PDF	40.3%
other	36.1%

Table 3 Results of Question 4

Q5: In which format do copy-editors indicate the changes to be made to articles?	
hard copy	5.6%
PDF with added embedded comments	38.97%
Word with comments and changes	26.4%
other	29.2%

Table 4 Results of Question 5

Question 4 and 5 request information regarding the formats in which the documents are passed between authors and copy-editors. Copy-editors need to be using a document format that they can make changes to. Question 4 can tell us what exact platforms are being used. In table 3, we can see that the majority of these journals are using Microsoft Word format for accepted version of manuscripts to copy-editors, and 40.3% of them turn the document into a PDF format then provide it to copy-editors. We can also find that only about 8% of journals[7] are providing LaTeX format to copy-editors. LaTeX files are often used to produce PDF files that are sent to copy-editors.

Table 4 shows about 39% of these copy-editors are using PDF with added embedded comments to indicate changes to be made to articles, while 26.4% of them are using Word. This question along with question 4 provides the information in terms of platforms: they can indicate the most possible places in which automations might happen in copy-editing process.

Q6: Who makes the copy-editing changes to the manuscript?	
author	40.9%
copy-editor	53%
typesetter	21.2%
other	22.7%

Table 5 Results of Question 6

Q7: In which format(s) do you provide the final revised versions of manuscripts to your typesetters?	
hard copy	4.6%
electronic-word	60%
electronic-PDF	38.5%
other	36.9%

Table 6 Results of Question 7

Q8: What text processing platform does typesetter use to create the final published article?	
Microsoft Word	21.5%
LaTeX	24.6%
don't know	53.8%
other	23.1%

Table 7 Results of Question 8

Question 6-8 are concerned with how copy-editors' and author' changes are provide to typesetter. From table 5, we can see that in over half of the journals, copy-editors make the change directly to manuscripts. These results tell us copy-editors could be likely to be most possible target audience to which we could introduce automations.

[6] The percentage was computed from the "Other" responses

[7] The percentage was computed from the "Other" responses

Table 6 indicates the format of final revised versions of manuscripts that is provided to typesetters. The majority of them are in Microsoft Word, while a smaller percentage of them are in PDF (38.5%) and a much smaller percentage in LaTeX (7%)[8]. The results of this question correlate with the results of question 8 provide us information that helps us to determine possible places where automations could happen in typesetting process.

Table 7 summarizes the text processing platform which is being used by typesetters. Although 53.8% of the participants do not know the details, Microsoft Word and LaTeX are the most two popular platforms.

Question 9 and 10 ask participants for other relevant comments.

Base on all these survey results, we can better understand the progress of production process. In the next section, we outline some possible production workflows in scholarly journal publishing.

4 Outcome of the Survey

As we have mentioned in section 2.2, automation would be mostly happen in copy-editing process (Dale, 1990). Since the survey results in table 2 show that 39 journals (46.9%) use in-house copy-editors to carry out copy-editing, we especially focus on these journals and outline 3 most common workflows of the production process[9].

Workflow 1 (11 journals out of 39):
1. Editor delivers a Word document to In-house Copy-editor
2. Copy-editor marks up the Word document with comments and changes
3. Copy-editor returns a corrected document to Editor
4. Editor returns the document to Author
5. Author approves changes and returns the document to Editor
6. Editor sends the Word document to Typesetter
7. Typesetter finalises the document using Word

Workflow 2 (8 journals out of 39):
1. Editor delivers a PDF document to In-house Copy-editor
2. Copy-editor marks up the PDF with comments and changes
3. Copy-editor sends the PDF to Editor
4. Editor returns the PDF document to Author
5. Author changes the Word document base on the comments in the PDF
6. Author returns a corrected Word document to Editor
7. Editor sends the Word document to Typesetter
8. Typesetter finalises the document using Word

Workflow 3 (7 journals out of 39):
1. Editor delivers a PDF document to In-house Copy-editor
2. Copy-editor sends comments and changes to Typesetter
3. Typesetter marks up the PDF document with comments and changes
4. Typesetter returns a corrected document to Editor
5. Editor returns the document to Author
6. Author approves changes and returns the document to Editor
7. Editor sends the PDF document to Typesetter
8. Typesetter finalises the document using LaTeX

After we have gathered adequate information of production process, it is possible for us to identify places in which automation might be introduced. In the next section, we suggest some kinds of automation which could benefit the publishing process.

5 Automation in the Publishing Process

In the editorial process, because the process of peer-review is the most frequent one, the target audience to be automated might be reviewers and authors. Reviewers might need an automation tool which could be fitted in to Word or LaTeX to help them mark up comments on the manuscripts rapidly and comfortably. Authors might need an automation tool which could help them easily correct their manuscripts.

In the production process, for the first workflow, the target audience might be copy-editors and typesetters. Cope-editors might need a tool which could be fitted into Word and easily mark

[8] The percentage was computed from the "Other" responses
[9] For other journals, their workflows of the production process will be presented in final report.

up and correct copy-editing errors. Typesetters might need a tool which could be fitted into Word as well but automates them with some typesetting tasks.

The possible automation is similar in the second workflow, whereas copy-editors might need additional tools that could be integrated with Acrobat and provide rapid mark up assistance for PDF files.

For the third workflow, the target audience might only be typesetters since they deal with most of the copy-editing and typesetting tasks. They might need an automation tool which could be fitted in Acrobat to help them with copy-editing, and another tool which could provide typesetting assistance with LaTeX.

There are also some automation tools that already exist in the market. For the editorial process, we can employ those manuscript management systems which are mentioned in section 2.1. For the production process, automated copyediting can be realized with some tools like Cadmus: Rapid Edit[10], Allen: TurnStyle[11] or Editors: Editors Toolkit[12]. Rapid Edit is developed by Cadmus Communications, and is customized to meet the specific standards of scholarly publishers. It can automate many repetitive style changes, corrections and tagging for copy-editors. Allen: TurnStyle is another smart tool which is also designed to automate repetitive and time-consuming copy-editing tasks. One of its best features is to cross-check in-text citations against reference list within minutes. Because these repetitive tasks such as reference formatting, standardizing abbreviations and eliminating extra spaces have been automated by the smart tools, copy-editors are able to focus on more substantive editorial tasks.

For typesetting, tools like Quark[13] or InDesign[14] are employed to help typesetters compose and prepare articles visually for print production. These tools are usually priced, e.g. InDesign costs US$699, and they are designed to process large volumes of manuscripts; so typesetting is typically outsourced.

[10] http://www.cadmus.com/products_and_services/

[11] http://www.allenpress.com/allen_press/gen/allen_press_generated_pages/Allen_TurnStyle_m101.html

[12] https://usd.swreg.org/soft_shop/47578/shopscr5.shtml

[13] http://www.quark.com/

[14] http://www.adobe.com/products/indesign/

6 Conclusion

Scholarly publication is a process involves tons of text processing tasks. Before we attempt to deploy any automated editorial assistance, we need to understand the variety of workflows and text processing platforms in the publishing process. In this paper, we discuss the workflows of editorial and production processes since they involve most of the text processing procedures. We investigate the production process in detail base on the results of the survey of journal publishing process. We also suggest some possible automation that might be introduced in both editorial and production process.

Finally, it should be noted that these efforts do not cover the whole picture of journal publishing process as the survey is restricted by its amount of questions and is limited to the computer science discipline: we can find that the results regarding typesetting are inadequate. In the future, a larger survey with more number of questions could be conducted to get more specific data of journal production process, especially for typesetting, from more disciplines.

References

Campbell, R., Meadows, J. & Page, G., 1997, *Journal Publishing*, Cambridge University Press, UK, Feb 1997 page 24.

Ciesielski, J., 2005, "How are electronic journals cooked", 11th Conference on Professional Information Resources, Prague, May 24-26, 2005, pp. 2-4.

Dale, R., 1990, "A rule-based approach to computer-assisted copy-editing", *Computer Assisted Language Learning*, vol.2, pp. 59-67.

Dale, R., 1997, "Computer assistance in text creation and editing", *Survey of the state of the art in human language technology*, Cambridge University Press, NY, pp. 235-237.

Mabe, M., 2003, "The growth and number of journals", *Serials*, vol.16, pp.191-197.

Steele, C., 2006, "Research with purpose", *The Australian Higher Education Section*, Sydney, page 42.