Reading Patterns and Formats of Academic Articles on the Web

Y. Rho and T.D. Gedeon

Abstract

Various formats are being used for Web-based academic articles such as conference papers and journal papers. We surveyed the formats being used and tried to identify reading activities and the proper formats by carrying out two online surveys: an email-based survey with an email-based questionnaire and a Web-based survey with a Web-based questionnaire.

The survey results show that readers overview Web-based academic articles from the screen, print them out and then read the printed articles. The results also show that the structural formats employed by most papers on the Web are against readers' preferences. The simple two-frame format was most preferred by 47% of the respondents as readers, but the cascade format of page windows was regarded as the worst by 65%. An interesting result is that 26% of the respondents selected as the worst style the paper-like format that is currently widely used for Web-based articles. Brief data sets and results are shown in this article.

In addition, the importance of examples embedded in the Web-based questionnaire was shown by two consecutive surveys.

Keywords: Web-based article; Reading patterns and formats; Remote preference gathering; Online survey

Introduction

Academic articles such as conference or journal papers can be regarded as a Web genre. There are lots of websites that contain academic articles. We have looked at some well-known sites including:

- ACM CHI97 (http://www.acm.org/sigchi/chi97/proceedings/paper/plp.htm);
- ACM/SIGCHI Bulletin (http://www.acm.org/sigchi/bulletin/1997.4/ross.html);
- ACM Transactions on Computer-Human Interaction (http://www.acm.org/pubs/contents/journals/tochi/1998-5/);
- Alertbox (http://www.useit.com/papers/webwriting/writing.html);
- International Journal of Human-Computer Studies (http://ijhcs.open.ac.uk);
- WWW6 Conference (http://www.scope.gmd.de/info/www6/technical/paper003/paper3.html); and
- WWW8 Conference (http://www8.org/fullpaper.html).

The article formats of these sites vary and reading activities for the articles have not been investigated well. Figure 1 shows one of the popular formats being used.

In order to identify the formats and activities, we applied two different online survey methods (see Root & Draper, 1983; Slauter, Haper & Norman, 1994; Hartson, Castillo, Kelso & Neale, 1996; Haper, Slauter & Norman, 1997; Perlman, 1997; Feinberg & Johnson, 1998): an email-based survey with a simple questionnaire and a Web-based survey with a Web-based questionnaire.

First survey: Email-based questionnaire

Purpose of the first survey

This survey was to see whether researchers find research articles from the Web and, if they do, how they read those Web-based articles?

Method

An email-based questionnaire of four questions was distributed to the fifty academic staff researchers and the eighty

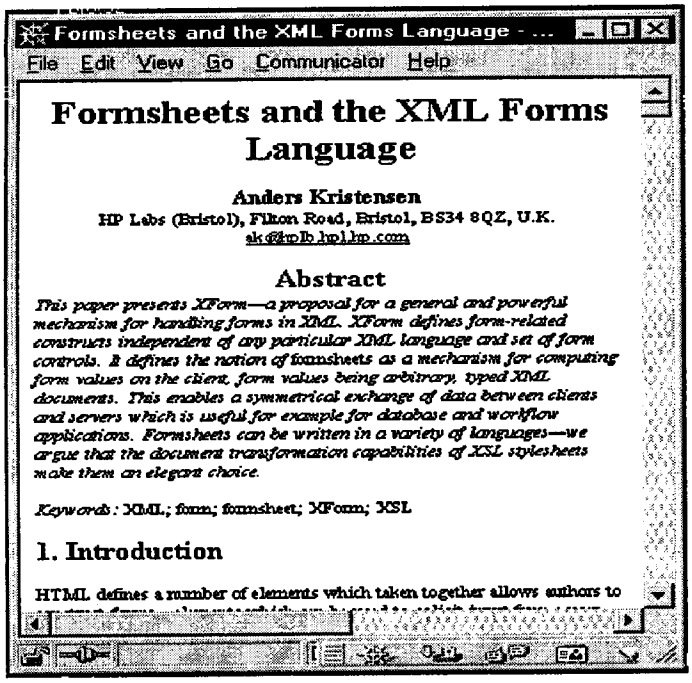

Figure 1: A popular format being widely used

(http://www8.org/w8-papers/1c-xml/formsheets/formsheets.html)

postgraduate research students in the School of Computer Science and Engineering at the University of New South Wales in Australia. No examples were given to subjects.

Results

We received twenty-three replies. The reply rate was 18%. Twenty-two (96%) answered that they find articles from the Web.

\	Details	1st choice		2nd choice	
Activity 1	Print and read	1	5%	1	5%
Activity 2	Read from the first screen, print and then read	0	0%	6	30%
Activity 3	Read concise parts, print and then read	14	64%	6	30%
Activity 4	Scan through, print and then read	7	32%	4	20%
Activity 5	Read from the screen	0	0%	3	15%
Others		0	0%	0	0%
N		22	100%	20	100%

Table 1: First survey result on reading activities with Web-based academic articles

Table 1 shows readers' usage patterns with Web-based academic articles. The majority of 64% selected Activity 3 for their first choice. Activity 4 was selected by 32% for their first choice. The selections for the other activities were not significant. An interesting result is that no one selected Activity 2 for the first choice. However, for the second choice, Activities 2 and 3 together recorded the highest selection occurrence of 30%.

Discussion

The Web is a resource that provides academic articles to researchers. Some sites such as digital libraries and Web journals are stable, but some sites such as conference sites are usually temporary. At the minimum, a stable source of academic articles must guarantee the existence of articles and should not change their information content over time.

The most typical usage pattern with Web-based articles is Activity 3. However, Activity 4 should not be ignored because it has been selected by 32% of the replies for the first choice. This resultant pattern is very similar to the result on usage patterns of paper-based academic journals in (Dillon, 1991a). It seems to be because academic articles also have a fine metastructure (see Dillon, 1991b, 1996) that leads to the reading activity of figure 2. The subjects of Activity 3 seem to want to view the concise parts from the first screen. On the other hand, the subjects of Activity 4 seem to focus much more on the content overview of an article than its interfaces.

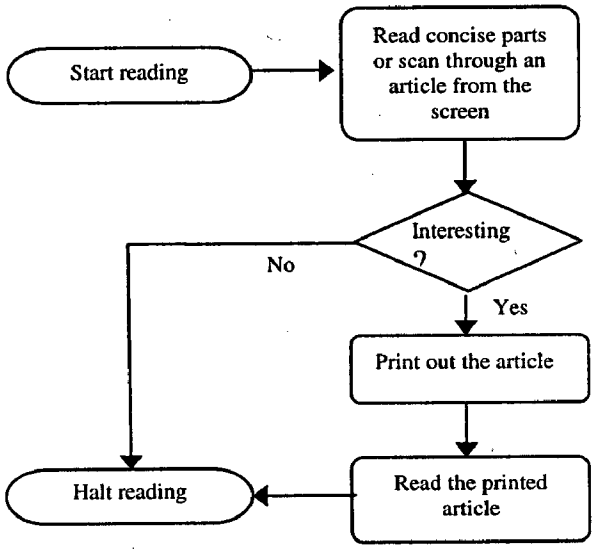

Figure 2: A typical reading flow from the first survey

In summary, readers find articles from the Web and get some ideas from the screen, print them out and read the printed articles, but seldom read them from the screen. Then, what formats do readers prefer and dislike? What will be the effect of interactive examples embedded in a questionnaire? The second survey was carried out to answer these questions.

Second survey: Web-based questionnaire

Purposes of the second survey

The second survey was to identify the format structure of Web-based academic articles which readers prefer and to discover the effects of examples in a questionnaire.

Methods

We prepared a Web-based three-part questionnaire with a total of eighteen questions in five groups (Rho, 1999). The first part of the questionnaire addresses window and screen size, leading the user through a sample configuration page to help avoid volume effects.

The second part consists of questions about three different features: overview types, windows layout and manipulation methods. The details of this part are not included in this paper.

The last part is about usage patterns and overall preferences in the different presentation format of Web-based articles. Each question has at least one corresponding example link. The same paper (Rho & Gedeon, 1998) was used for all examples in the questionnaire.

The survey was announced via email with the questionnaire URL to researchers in information technology. They were postgraduate research students, research staff and academic staff. Neither undergraduate nor postgraduate coursework students were included. The number of candidate subjects was one hundred and fifty.

Results

We received thirty-four replies: thirty from researchers in the information technology area and the remaining four from researchers in the education and engineering areas. The reply rate was 23%, thirty-four out of one hundred and fifty. Most of them (82%) used 17" monitors to fill out the questionnaire.

Table 2 shows the reading activity patterns that our thirty-four respondents selected for their first and second choices. Like the first survey result, Activity 3 was the most typical reading activity that 55% of our subjects selected, but Activity 2 and 4 were selected by 18% each. In the first survey, the ratios were 0% for Activity 2 and 32% for Activity 4, which are very different from the second survey results. Again, those two activities share the same ratio of 29% in the second choice.

\	Details	1st choice		2nd choice	
Activity 1	Print and read	2	6%	2	6%
Activity 2	Read from the first screen, print and then read	6	18%	10	29%
Activity 3	Read concise parts, print and then read	19	56%	8	24%
Activity 4	Scan through, print and then read	6	18%	10	29%
Activity 5	Read from the screen	1	3%	4	12%
Others		0	0%	0	0%
N		34	100%	34	100%

Table 2: Reading activities

Table 3 shows that 47 %, sixteen out of thirty-four subjects, selected the two-frame format for the best and 35% selected

\		Best(1st)		...	Worst(5th)	
Format 1	Paper-like	2	6%		9	26%
Format 2	Paper-like with TOC*	12	35%		0	0%
Format 3	Two frames	16	47%		0	0%
Format 4	Slides	2	6%		1	3%
Format 5	Cascades	2	6%		22	65%
N/A		0	0%		2	6%
N		34			34	

TOC*: Table of Contents

Table 3: Preferences in overall web article formats

the paper-like-with-TOC format for the best. Meanwhile, 65% selected the cascade format as the worst. It is interesting that 26% of the respondents selected the paper-like format as the worst. The paper-like-with-TOC format and the two-frame format recorded none for the worst.

Table 4 shows the correlation between the first-choice activity selections and the first-choice format selections. For example, the underlined shaded cells show the format distribution over for people who prefer Activity 3. The Format 3 contribution to Activity 3 is highest at 47% and Format 2 is next at 32%. Similarly, the shaded column corresponds to Activity distribution over Format 3. Activity 3 is 56%, which is followed by Activities 4 and 2. Activity 3 is most correlated with Format 3 and vice versa. Activities 4 and 2 are also most correlated with Format 3, with the contributions of 67% and 50% respectively.

Discussion

The most typical reading activity with a Web-based academic article is for readers to read some concise parts of an article, print out the article if they are interested in it, and then read the printed article. Both surveys produced this result.

A big difference between the two surveys on the reading patterns happened at Activities 2 and 4. In the first survey, the portion for Activity 4 was 32% but 0% for Activity 2. Then, in the second survey, the portion for Activity 4 decreased by 14% to 18% but the portion for Activity 2 increased by 18%. So the portions are even.

Why? There was no significant difference in the distribution list and the question for this topic. The difference seems to have been from the existence of examples that the subjects could experience. The first questionnaire was based on email with no examples to try. Meanwhile, the second questionnaire was distributed on the Web with many examples. Subjects had many chances to try different formats of a Web-based article to answer the preceding questions. The subjects seem to have been aware of the importance of the first screen, but not their subsequent behaviour.

\Formats Activities	Format 1			Format 2			Format 3			Format 4			Format 5			N(Activity)	
Activity 1	0			2	100%	17%	0		0%	0			0			2	6%
Activity 2	1	17%	50%	1	17%	8%	3	50%	19%	0			1	17%	50%	6	18%
Activity 3	1	5%	50%	6	32%	50%	9	47%	56%	2	11%	100%	1	5%	50%	19	56%
Activity 4	0			2	33%	17%	4	67%	25%	0			0			6	18%
Activity 5	0			1	100%	8%	0		0%	0			0			1	3%
N(Format)	2	6%		12	35%		16	47%		2	6%		2	6%		34	100%

Table 4: Correlation between reading activities and formats for the two first choices

The two-frame format (Figure 4) was most preferred by respondents. In the most frequent reading activities, a common bridge from the screen to paper is printing. Article interfaces for readers have to be able to support two different media: the Web and paper. The majority of respondents selected the two-frame format as the best one for the purpose, which is against the general idea that using frames is not good. The paper-like-with-TOC format can be an alternative choice. The cascade format was worst because of its complexity on the screen, but no one selected either the two-frame format or the paper-like-with-TOC format for the worst.

Activities and formats showed correlation: Activities 3, 2 and 4 most match the two-frame format (Table 4).

The early part of the reading process, which is to overview an article, happens on the screen when reading a Web-based academic article (Figure 3). The survey results show that its early parts, overviewing and printing, are most likely to be supported by the two-frame format that ensures the quality of the paper-based article format when printed.

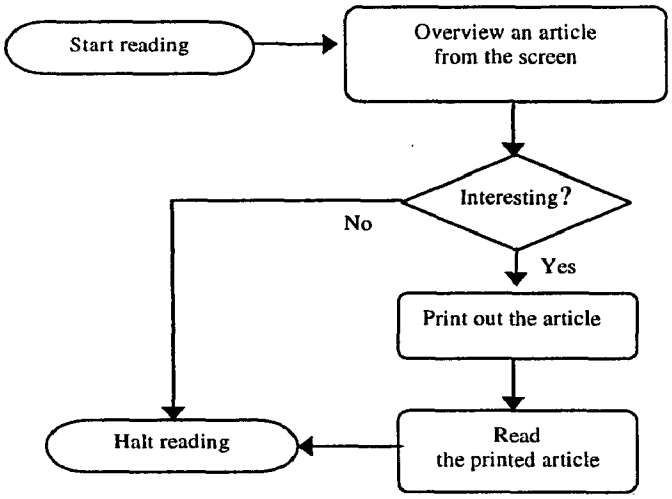

Figure 3: A typical reading flow from the second survey

Conclusions

We identified the typical reading activity pattern and the proper structural format for Web-based academic articles, based on end users' selections from two online surveys. We additionally examined the activity-format correlation.

The results show:

- Most readers overview Web-based articles from the screen, print them out and then read the printed articles;
- The two-frame format seems to best support different reading activities in general, followed by the paper-like-with-TOC format; and
- Embedded examples in a questionnaire are likely to affect survey participants.

This research focused on only the structural format of Web-based academic articles from the viewpoints of information, interfaces and interactions. Further research on other features considering reading activities should be done.

References

Dillon, A. (1991a). "New Technology and the Reading Process," *Computers in Libraries*, July 1991, 23-26.

Dillon, A. (1991b). "Readers' models of text structures: The case of academic articles," *International Journal of Man-Machine Studies*, 35, 913-925.

Dillon, A. (1996). " Expertise and the Perception of Shape in Information," *Journal of the American Society for Information Science*, 47(10), 786-788.

Feinberg, S. & Johnson, P. (1998). "Designing and Developing Surveys on WWW Sites,"*Proceedings of Computer Documentation '98.*

Harper, B., Slaughter, L., Norman, K. (1997). "Questionnaire administration via the WWW: A validation & reliability study for a user satisfaction questionnaire," *Proceedings of Webnet'97.* http://www.lap.umd.edu/QUIS/webnet.pdf

Hartson, R., Castillo, J., Kelso, J. & Neale, W. (1996). "Remote Evaluation: The Network as an Extension of the

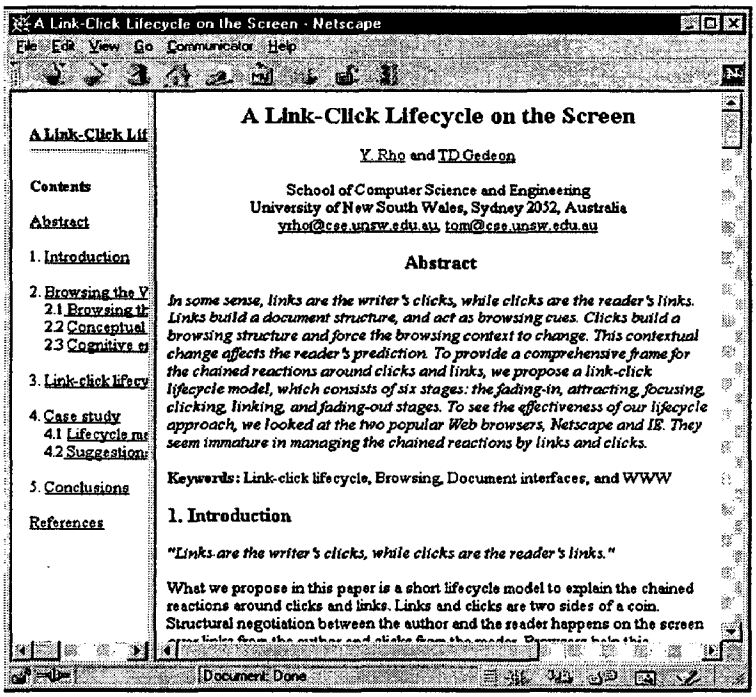

Figure 4: The two-frame format given as an example (Rho & Gedeon, 1998)

Usability Laboratory," *CHI'96 Electronic Proceedings.* http://www.acm.org/sigchi/chi96/proceedings/papers/Hartson/hrh_txt.htm

Slaughter, L., Harper, B. & Norman, K. (1994). "Assessing The Equivalence Of The Paper And On-line Formats Of The Quis 5.5," *Proceedings of the 2nd Annual Mid-Atlantic Human Factors Conference*, Washington, D.C, 87-91. http://lap.umd.edu/LAPFolder/Papers/SHN.html

Rho, Y. & Gedeon, TD. (1998). "A Link-Click Lifecycle on the screen," *Proceedings of Asia-Pacific Web Conference: Web Technologies and Applications*, Beijing, 307-312. http://www.cse.unsw.edu.au/~yrho/Publications/apweb98/apweb98_C-tiled_main.htm

Rho, Y. (1999). "Questionnaire on Reader's Preferences in Web articles." http://www.cse.unsw.edu.au/~yrho/WebPaper/ Question.htm

Root, R. W. & Draper, S. (1983). "Questionnaires as a software evaluation tool," *Proceedings of Computer Human Interactions*, 83-87.

About the Authors

Mr. Young Rho is a Ph.D. student in the School of Computer Science and Engineering at the University of New South Wales in Sydney. His research interests are document-oriented interfaces and evolutionary Web design methods.

Dr. Tom Gedeon is a Professor in the School of Information Technology at Murdoch University and a General Chair of the 6th International Conference on Neural Information Processing. His research interests are information retrieval and index generation, learning algorithms, and neural networks and fuzzy logic.

Authors' Addresses

Young Rho
School of Computer Science & Engineering
University of New South Wales
Sydney, NSW 2052, AUSTRALIA
yrho@cse.unsw.edu.au

Tom Gedeon
School of Information Technology
Murdoch University
Murdoch, WA 6150, AUSTRALIA
tom@it.murdoch.edu.au