

An exploration of ebook selection behavior in academic library collections

Dana McKay¹, Annika Hinze², Ralf Heese³, Nicholas Vanderschantz²,
Claire Timpany², Sally Jo Cunningham²

¹Library, Swinburne Institute, PO Box 218 John Street, Hawthorn VIC 3122, Australia
dmckay@swin.edu.au

²Dept. of Computer Science, University of Waikato, Private Bag 3105, Hamilton New Zealand
{hinze, vtwoz, ctimpany sallyjo}@waikato.ac.nz

³Institute of Computer Science, Freie Universität Berlin, Königin-Luise-Straße 24/26, Berlin
heese@inf.fu-berlin.de

Abstract. Academic libraries have offered ebooks for some time, however little is known about how readers interact with them while making relevance decisions. In this paper we seek to address that gap by analyzing ebook transaction logs for books in a university library.

Keywords: ebooks, log analysis, book selection, HCI, information behavior

1 Introduction

Consider the process of borrowing a book from a library (digital or physical): the reader searches or browses the collection to identify candidate books that are potentially relevant; they each candidate book to assess its actual relevance; and they borrow those books deemed relevant to explore them more depth at a later date. Of course, finding a useful book is rarely so straightforward: these stages may be sequential or interleaved.

In this paper, we consider a specific aspect of this process: the examination of a candidate book after the candidate has been identified in the library collection; we focus specifically on ebooks. How a reader explores physical books when making relevance decisions—which parts of books are viewed, how quickly, in what order—has been relatively neglected [8], likely because such a study in a physical domain would be intrusive and ‘creepy’ [16]. To our knowledge this aspect of ebook selection has also not been explored, surprisingly, given that the availability of ebook transaction logs allows such a study to be conducted post hoc so as to avoid disturbing the experience of those using an ebook collection. This paper describes exactly such a study based on the transaction logs of a university library ebook collection.

In Section 2 of this paper we explore the previous research on ebook usage and book selection (both physical and electronic); in Section 3 we describe the ebook collection on which our study is based and gives details of the log sample we analyzed. Section 4 presents the results of our analysis, including which parts of the

books readers examined, and readers' interaction patterns with the ebooks; Section 5 discusses the relationship of these results to the existing understanding of document navigation and selection; and Section 6 presents our conclusions.

2 Related work

The work related to our study falls into two categories: the use and usability of ebooks (Section 2.1) and the literature on book selection (Section 2.2).

2.1 Use and usability of ebooks

The use of ebooks for recreational reading is increasing: in 2011 purchases of ebooks on Amazon surpassed those of print books [6]. In the academic sphere—where our study takes place—some disciplines demonstrated higher use of ebooks than print books since as early as 2002 [4], and a slight institution-wide predominance was seen in one study in 2004 [11]. Nearly all studies of ebook use, though, show that uptake varies by discipline (it is more common in technical disciplines) and reader circumstances (for example reader location).

The online nature of ebooks accounts for many of the pros and cons given by readers: readers liked the searchability and currency of ebooks, and the ability to access them anywhere [9, 14], however they were frustrated by restrictions on printing and copying, and by DRM [22]. Users' other frustrations include poor ebook annotation capabilities [16] and in-book browsing facilities [2, 10]. Ebooks were perceived to be more comfortable read on dedicated reading devices than on computer screens [9], but even these devices have problems [15, 19], and a study of students given Amazon Kindles to trial showed the majority did not use them for study-related reading [25].

Despite these problems, however, the largest and most recent survey of academic ebook usage showed over half of all respondents had used at least one ebook [9].

2.2 Book selection

Rowlands *et al.* noted in 2007 [21] that book selection was a surprisingly unstudied part of the book use process, an assessment that remains true today.

Studies of children's book selection practices are more common than those of adults: Reutzel and Gali [20] followed a number of children through the process of checking out library books, and noted that children were more likely to choose books from eye-level, tended to make judgments based on color, and that while they occasionally flipped through books, they rarely made decisions on the basis of content. Moore [18] witnessed children choosing the first book on any shelf related to their topic, as opposed to browsing to select the most appropriate volume for their needs.

Stelmaszweska *et al.*'s work [23] following computer scientists is the earliest published work based on observation of adults selecting books in a library. This study demonstrated that while these readers did use book content to guide their decision to some extent, they also looked at the book covers to determine books' ages and as-

sessed the amount of dust on books to determine whether they had been recently used. In a physical library, of course, availability is also an issue. Our own previous work on adults in an academic found many of the same things, for example that covers are important and readers make decisions based on perceived book age [8]. Buchanan's bookshop study [3] provides some examples of people choosing between books, and notes that most book shoppers did not, in fact, open the books when making their decisions; however these examples all describe shoppers buying for someone else.

At a more fine-grained level, Stieve *et al.* [24] asked university students to choose between two similar books to meet a pre-defined information need. In this study researchers observed and students reported using the Table of Contents (ToC) and the organization of books to make decisions. The importance of ToC, especially in ebooks, is reinforced by other studies [2, 10]. This finding is also reflected in our companion work on book selection practices in physical libraries [8].

Despite the broad number of studies, there is little work on the actual process readers use to select books of interest or reject those that are not interesting, nor on the data they use to make these decisions. From the studies described above we might guess that choices are based on some combination of book-cover, non-metadata information (such as dust) and the table of contents: our study investigates this question further for academic ebooks.

3 Methodology

Our study used data gathered at Swinburne University of Technology in Melbourne, Australia. Swinburne is a small, dual-sector (university and polytechnic), research-active institution. It has approximately 27,000 students, 770 research postgraduates, and 1270 academic staff. This section describes the ebook collection the data came from and the method used in this study.

3.1 The ebook collection

The ebooks studied in this paper are provided by EBL (<http://www.ebllib.com>), a large ebook provider in Australia. Swinburne makes nearly 20,000 books from the EBL collection available via the library catalogue; some of these owned by the library, but the majority are available on an ad-hoc basis. The difference between owned and un-owned ebooks is transparent to readers; Swinburne uses a patron-driven demand model [7] for ebooks. EBL allows readers five minutes of browsing within books that are not owned by the institution, and ten minutes in books that are. If readers wish to continue to use books beyond this time, or if they wish to copy or print any content the system presents a dialog box inviting users to create a loan; to continue using the book users must click 'yes' on this dialog box, thus providing an affirmative expression of interest. This is the only means by which readers borrow books—they cannot actively decide to check out an ebook and click a button, for example.

EBL books are presented in a web-browser via the EBL interface (Fig. 1); they are not downloadable. The interface allows navigation through ebooks in a variety of

ways: the use of a right-hand scrollbar, a left-hand navigation menu based on the table of contents, and a paging navigation interface above the book.



Fig 1. The EBL interface

3.2 Ebook log analysis

The data on which this study was based was collected over the period 1 September to 30 September 2011; this period falls completely during term time at Swinburne, meaning the data is likely to represent a range of academic users, rather than just academic staff and research students. During the study period 9506 people accessed ebooks, of whom 3799 went on to create a loan.

EBL logs anonymised use by individual readers, and captures quite detailed statistics including which books have been used, which pages were examined in each book, the number of minutes spent looking at a book, whether the book was owned by Swinburne, and bibliographic details about the book. When scrolling, only those pages on which a user pauses are logged in the usage stats; pages they scroll past are not.

As we were interested in how readers select (or do not select) ebooks, we created two samples: one of data about books readers went on to borrow, and one of books that readers only browsed. Note that a book not having been borrowed does not mean it was not relevant to the reader's information need, for example they may have found the information they were looking for during the available browsing time or they may have found another more useful book.

The sample on which our study is based is data from 100 randomly selected browsed books, and 100 loaned books, each accessed by a different user. All data is from the period while readers were making a decision, i.e. before a loan was created, and includes all the pages readers viewed to a maximum of 19 pages. We classified the content of these pages using a classification scheme based partly on Reutzel and Gali's observations of children [20], and partly on McKay's earlier work on ebooks [14]. We then examined these results for patterns, and compared loaned books to browsed books.

4 Results

The results of this study can be broadly classified into three categories: the book features readers viewed, the length of time readers spent with books, and how readers examined books and their features.

4.1 Parts of books readers viewed

Our analysis discovered that the five most commonly viewed parts of the book were front matter, chapter headings, table of contents, the first page of content, and the introduction (see Fig. 2). Most commonly used was the front matter; 93% of loaned books and 89% of browsed books showed some use of this content, (this may be an artifact of the EBL system, which displays front matter first by default). 52% (loaned books) and 56% (browsed books) of readers viewed chapter headings. Surprisingly, given that an interactive version of the ToC is available to readers in the left-hand navigation, 41% and 43% of readers viewed the non-interactive printed in-book ToC at some stage, perhaps to investigate page numbers (which are not usually present in the left-hand ToC). Very few readers (2 who borrowed books and 1 who merely browsed) used the index, 3 and 1 respectively used the bibliography, and only one reader—who borrowed the book—viewed the conclusions section. Readers also viewed the introduction (25% loaned / 24% browsed), first page of content (23% / 33%), blank pages (33% / 21%), images (22% / 21%) and tables (13% / 12%).

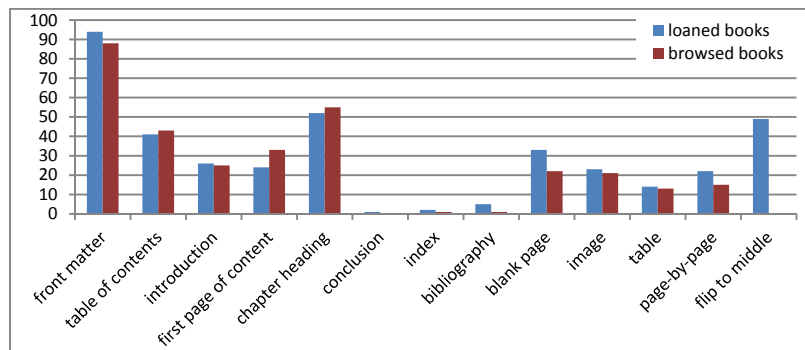


Fig 2. Parts of books viewed by readers (loaned and browsed books)

4.2 First three items viewed

Almost all readers reviewed the front matter of a book before initiating a more thorough investigation of the book, however this behavior is not predictive of the use of other in-book metadata (for example ToC). Figs. 3 and 4 show the most commonly viewed first, second and third items. Readers typically began moving through a book in a page-by-page manner, either by using the top navigation buttons or by scrolling (see Fig. 1). In 160 books (76 browsed, 84 loaned) this sequential examination of the front matter lasted three pages or more.

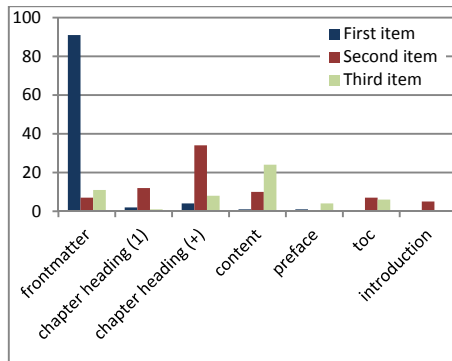


Fig 3. 1st three parts viewed of 'loaned' books

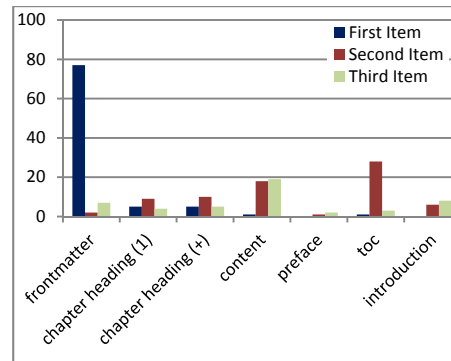


Fig 4. 1st three parts viewed of 'browsed' books

4.3 Time taken to make a decision

Readers typically either made a decision about a book very quickly or used all of their available browsing time (5 minutes for unowned ebooks, 10 minutes for owned ebooks). In 36% of browsed books and 26% of loaned books readers made a decision within the first minute, rising to 53% and 40% respectively at two minutes. Conversely in 30% of read books and 13% of browsed books users browsed for the available browsing time before making a decision. In all cases—browsed and loaned, owned and unowned books—the data forms a u-shaped curve.

Readers viewed between 0.3 and 24 pages per minute in our sample, with a median rate of 5.3 pages per minute. Readers viewed pages more quickly in browsed books (median 6.9 pages per minute, mean 7.1) than in loaned books (median 4.5, mean 5.7), a difference that is significant at $p=0.017$. This difference suggests that readers read more carefully or thoroughly in books they go on to borrow than those they do not, a finding that echoes McKay's earlier work in this area [14].

4.4 Reading sequences of pages

Readers were seen to move page by page through the book (21% loaned / 14% browsed), flip to the middle of a section within the content (49% / 29%), and directly navigate to a chapter heading (51% / 55%). Based on the page visited before and after a page being coded we are able to make assumptions about the ways in which users navigated books. Readers appeared to use the left-hand ToC navigation more often than they entered a page number into the top navigation: 63 of 200 readers in our study apparently used the left-hand navigation while only 14 of 200 entered a specific page number in the top navigation section.

The majority of readers viewed sequences of 3 or more consecutive pages at some stage during their interactions. Borrowers were more likely than browsers to look at more one sequence of 3 or more consecutive pages. Of those that borrowed the books, 35 looked at 1 sequence of three or more pages, 37 looked at 2 sequences, 17 at 3

sequences and only 6 readers interacted with 4 sequences. Of those who only browsed, 44 looked at 1 sequence, 29 looked at 2 sequences and 12 looked at 3 sequences. No reader viewed more than four sequences, and only those who borrowed books viewed more than three.

Many of the readers who looked at just one sequence of consecutive pages did so at the beginning of their interaction. After the first three pages, most readers (60% in loaned books, 43% in browsed books) navigated either to another part of the book (50% loaned, 30% browsed) or away from the book entirely, ending their exploration at this point (10% loaned, 13% browsed). A very small number of readers (4% loaned, 1% browsed) viewed all 19 sample pages consecutively.

4.5 Navigation within a book

We looked at the sequences of pages accessed in order to explore the ways in which readers navigated books. Readers generally began by paging through a book, either by scrolling or using the paging navigation buttons. (see Fig. 1). This behaviour typically continued through the first three pages of the front matter of a book before readers discovered the left-hand ToC (see Fig. 1), and used this to continue their search. This pattern was seen in 126 of 200 readers. Readers were seen to move page-by-page through books (21% in loaned books, 14% in browsed books) and flip to the middle of a section (49% / 29%) and directly navigate to a chapter heading (51% / 55%).

4.6 Interaction patterns

To investigate reading patterns we encoded readers' moves between pages as either skips (1 page) or jumps (more than 2 pages, and usually more than 10 pages) forward or back. From this encoding, we identified three patterns: linear progression, contextual confirmation, and exploratory assessment, described further below.

Linear progression (shown in Fig. 5) involves readers paging through the initial parts of the book (we assume using the paging navigation buttons or scrolling), before using the left-hand ToC navigation to jump forward in the book to the start of a chapter. This pattern reflects behaviour observed in our earlier work in physical libraries [8], where readers flipped through books and reading parts of relevant chapters.

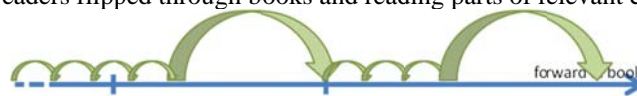


Fig 5. Linear progression (pattern A).

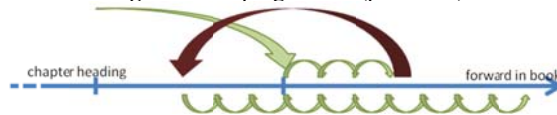


Fig 6. Contextual confirmation (pattern B).

The second interaction pattern (Fig. 6) represents a readers jumping to the first page of a chapter (most likely using the left hand ToC navigation) before paging forward two to three pages then jumping backward in the book to the final few pages of

the previous chapter. From there they read to their previous reading position and then read further forward in the book. We assume that the reader is reviewing the end of the previous chapter for context to assist with the reading within the current chapter. This interaction pattern was not observed in readers in physical stacks [8].

The third interaction pattern (Fig. 7) shows the reader jumping back and forward throughout the book, seldom looking at more than one page. This interaction is most likely a result of scrolling (only those pages that completely load are noted in logs), but may also be a result of searching or entering page numbers in the paging navigation. Without knowing which method of interaction readers were using to create this pattern it is difficult to speculate on their intentions.

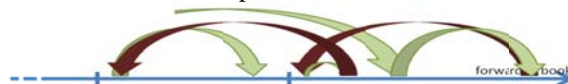


Fig. 7. Exploratory assessment (pattern C).

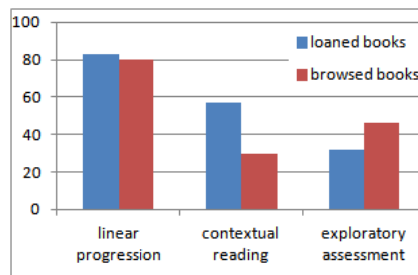


Fig. 8. Distribution of interaction patterns

The patterns were not evenly distributed (see Fig. 8) Most readers used pattern A, linear progression, at some point during their reading—both those who borrowed (83%) and those who browsed (80%) books. Pattern B, contextual confirmation, was the second most frequently seen pattern in loaned books but least likely to be seen in browsed books (30%). Pattern C, exploratory assessment, was more often evident in browsed books (46%) than in loaned books (32%). These patterns were relatively clearly distinguishable in reading behavior; most readers used only one or (more commonly) two of the patterns; use of all three was rare (see Fig. 9)

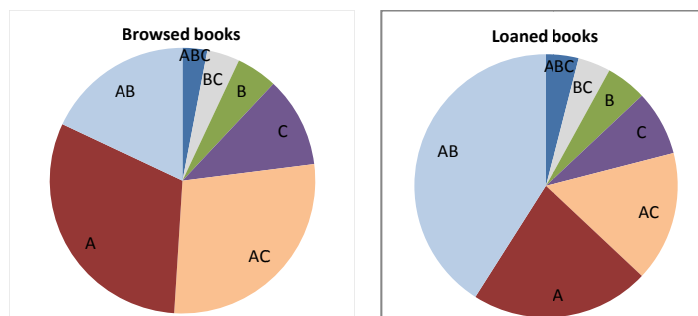


Fig. 9. Proportion of reading patterns used by readers.

Pattern A (linear progression) was most likely to be the only pattern used; patterns B (contextual confirmation) and C (exploratory assessment) were most commonly used in conjunction with other interaction patterns. In loaned books, a stronger focus on extended reading (pattern AB) is evident even in the first 20 pages, whereas browsed-only books are often approached in a more exploratory manner (AC).

5 Discussion

There are two major areas of work related to our findings: document navigation, and document selection literature.

5.1 Document navigation

The navigation patterns seen in Section 4.6 are very interesting: they show a considerable divergence from our work in physical environments, where readers do not typically flip backwards, and especially not in large chunks.

What these patterns do show is consistency with the interaction patterns discovered in the same collection in [14], and with the patterns seen in [13], which examines digital reading of journal articles. The differences in interaction between the ebook collection examined here and our earlier work in physical environments [8], and the backward flipping seen in common with digital article reading in [13], tends to suggest that there is something about digital environments that facilitates or encourages significant backwards movement in text. This conclusion may be supported by comparing [17] which shows some small backward movement in print periodicals to [12] which shows frequent large jumps backward in electronic articles.

If readers are only flipping backward because they can, and not because they must (either to gain context [17], or, as suggested by some commentators—for example [5]—because it is not possible to read deeply online), then digital books are supporting users somehow in a way in which physical books do not. If, conversely, users flip backwards in digital books because they are in some way lost (and users do frequently get lost in digital books, as seen in [10, 15]), then DLs must understand and address the cause of this disorientation to best support users' reading activities. It is as yet an open question whether the large jumps back seen in this and other studies on ebooks are due to disorientation or inclination; further study is needed to clarify this issue.

5.2 Document selection

The common investigation of front matter seen in our study is unusual among book selection studies. While our own earlier work saw some readers checking for edition information [8], and a charitable interpretation of this behavior would assume that readers were using this material—particularly the copyright page—it is unlikely that this is the case. Buchanan and McKay show that the majority of those seeking books even in high end bookshops have a limited knowledge of the kinds of metadata (such as publisher) contained in front matter [3], and [1] and [26] demonstrate that both

academics and university students struggle to create citations (using the kind of information found in front matter). Malama *et al.* found that, at least for fiction books, many readers felt that ebooks had too much introductory material and would have preferred to go straight to the text [15]. It is almost certain, then, that this frequent use of front matter is an artifact of the EBL system (which defaults to readers seeing front matter first), though this front-dominance in the digital realm has been seen in one other study comparing electronic articles to print [12]. It is likely that some other information than front matter is more useful in decision-making, and so DL designers should provide this information to readers first and foremost. What information that might be, however, remains an open question, though some of the major candidates are discussed below.

The EBL system provides readers with instant access to a navigation panel to the left of the book that, for most books, displays some kind of ToC. The ToC is seen to be useful in decision-making in print books [2, 8, 10, 23, 25], print articles [17] and ebooks [2, 10, 15]. Readers in our study used the interactive ToC in the left-hand panel relatively frequently (62% of all readers), suggesting that many at the very least saw this panel. It is surprising, then, how many readers (~40%) nonetheless viewed the in-book version of the ToC, which is not interactive. Either these two groups are mutually exclusive, or the print ToC offers something that the interactive one does not: one possible candidate for this information is page numbers, which are usually available only in the in-book version of the ToC. Page numbers could conceivably be used to determine how much of a book is dedicated to a topic of interest, and therefore should probably be available from any version of a ToC.

Earlier studies show the frequent use of index for information seeking and decision-making within books in both physical [8] and digital formats [2]: a sharp contrast to the data seen here where only a few readers accessed the index. The reason for this discrepancy remains to be discovered.

Images are seen as a source of decision-making data in our own earlier work on print books [8], in children's book decision making [20], and in studies of triage both in physical periodical materials [17] and electronic articles [13]. The lack of use of images in the work presented here is likely due to the long load time of many pages during scrolling, meaning that images are not visible to readers flipping past them in EBL in the same way that they are in print books or pre-loaded electronic documents.

Finally, that so few readers consulted the conclusions section of books in decision-making, even though this section was seen as very important for triaging electronic articles [12, 13] and it is readily available using the left-hand navigation, is perplexing. It may be that conclusions are not as useful in decision-making in long-form documents such as books as they are in short-form documents, or perhaps navigating to conclusions over a long space in a digital environment is prohibitively difficult. This remains a question for future work, but has profound implications for DL design.

6 Conclusions and future work

This study raises nearly as many questions as it answers in terms of how readers make decisions about ebooks: ToC is frequently used, but the reasons for the tensions between print and interactive versions of the ToC are not clear. Conclusions and index seem to be less useful than in previous studies, though the reasons are for this are not readily apparent. Front matter is frequently used, though it looks as though this may be an artifact of the EBL system: which information would be most useful presented first is a subject for future study. Finally, images are infrequently used in EBL, almost certainly due to the system failing to present them to readers; DL designers would do well to incorporate image presentation as a key feature of ebook display environments.

We do not yet have a clear understanding of whether readers' needs are being met by the current EBL interface. The current interface provides three main methods of navigation for the reader, the left-hand ToC, the top navigation and scrolling; however, there may be different interaction methods that would better support the sampling activities readers want to conduct. This research could be extended to gain understanding of whether current interaction tools adequately facilitate the identified reading patterns and whether other interface features would better support readers' sampling activities.

Acknowledgements. We acknowledge with thanks the kind support of Alison Morin and EBL in facilitating access to the EBL collection. Thanks also to Tony Davies, Nguyen Ly and Fiona Campbell for their technical support for this research.

References

1. Aronsky, D., J. Ransom, *et al.* (2005). Accuracy of References in Five Biomedical Informatics Journals. *J Am Med Inform Assn* **12**(2): 225-228.
2. Berg, S. A., K. Hoffmann, *et al.* (2010). Not on the Same Page: Undergraduates' Information Retrieval in Electronic and Print Books. *J Acad Libr* **36**(6): 518-525.
3. Buchanan, G. and D. McKay (2011). In the Bookshop: Examining Popular Search Strategies. *JCDL 2011*. Ottawa, Canada, ACM.
4. Christianson, M. and M. Aucoin (2005). Electronic or print books: Which are used? *Libr Collect Acquis* **29**(1): 71-81.
5. Cull, B. W. (2011). Reading revolutions: online digital text and implications for reading in academia. *First Monday* **16**(6).
6. Hamblen (2011). Amazon: E-books now outsell print books. *Computerworld* http://www.computerworld.com/s/article/9216869/Amazon_E_books_now_outsell_print_books 2011, last accessed 11 April 2012.
7. Hardy, G. and T. Davies (2007). Letting the patrons choose: using EBL as a method for unmediated acquisition of ebook materials. *Information Online 2007*: Sydney, Australia, ALIA.
8. Hinze, A., McKay, D., Timpany, C., Vanderschantz, N., and Cunningham, S.J. (2012) Selection behaviour in the physical library. *JCDL 2012*, Baltimore, MD. To appear.

9. Li, C., F. Poe, Potter, M, Quigley, B and Wilson, J. (2011). *UC Libraries Academic e-Book Usage Survey*. Oakland, CA, USA, University of California, Springer, California Digital Libraries.
10. Liesaputra, V. and I. H. Witten (2008). Seeking information in realistic books: a user study. *JCDL '08*. Pittsburgh PA, PA, USA, ACM: 29-38.
11. Littman, J. and L. Connaway (2004). A Circulation Analysis of Print Books and E-Books in an Academic Research Library. *Library Resour Tech Ser* **48**: 256-262.
12. Loizides, F. (2007). Investigating Document Triage on Paper and Electronic Media. *ECDL 2007*, Springer Berlin / Heidelberg. **4675**: 416-427.
13. Loizides, F. and G. Buchanan (2009). An Empirical Study of User Navigation during Document Triage. *ECDL 2009*, Springer Berlin / Heidelberg. **5714**: 138-149.
14. McKay, D. (2011). A jump to the left (and then a step to the right): reading practices within academic ebooks. *OZCHI 2011*. Canberra, Australia, ACM: 202-210.
15. Malama, C., M. Landoni, et al. (2004). Fiction Electronic Books: A Usability Study. *ECDL '04*, Springer Berlin / Heidelberg. **3232**: 69-79.
16. Marshall, C. C. (2010). *Reading and Writing the Electronic Book*. Chapel Hill, NC USA, Morgan & Claypool.
17. Marshall, C. C. and S. Bly (2005). Turning the page on navigation. *JCDL'05*. Denver, CO, USA, ACM: 225-234.
18. Moore, P. (1995). Information Problem Solving: A Wider View of Library Skills. *Contemp Educ Psychol* **20**(1): 1-31.
19. Pearson, J., G. Buchanan, et al. (2010). HCI design principles for ereaders. *Booksonline '10*. Toronto, ON, Canada, ACM.
20. Reutzell, D. R. and K. Gali (1998). The art of children's book selection: A labyrinth unexplored. *Reading Psychology* **19**(1): 3-50.
21. Rowlands, I., D. Nicholas, et al. (2007). What do faculty and students really think about e-books? *Aslib Proc* **59**(6): 489-511.
22. Shelburne, W. A. (2009). E-book usage in an academic library: User attitudes and behaviors. *Libr Collect Acquis* **33**(2-3): 59-72.
23. Stelmaszewska, H. and A. Blandford (2004). From physical to digital: a case study of computer scientists' behaviour in physical libraries. *JoDL* **4**(2): 82-92.
24. Stieve, T. and D. Schoen (2006). Undergraduate Students' Book Selection: A Study of Factors in the Decision-Making Process. *J Acad Libr* **32**(6): 599-608.
25. Thayer, A., C. P. Lee, et al. (2011). The imposition and superimposition of digital reading technology: the academic potential of e-readers. *CHI 2011*. Vancouver, BC, Canada
26. Yu, F., J. Sullivan, et al. (2006). What Can Students' Bibliographies Tell Us?- Evidence Based Information Skills Teaching for Engineering Students. *Evidence Based Library and Information Practice* **1**(2).