

In Search of a Good Novel Examining Results Matter

Suvi Oksanen

School of Information Sciences
University of Tampere
33014 University of Tampere, Finland
Suvi.Oksanen@uta.fi

Pertti Vakkari

School of Information Sciences
University of Tampere
33014 University of Tampere, Finland
Pertti.Vakkari@uta.fi

ABSTRACT

We studied how an enriched public library catalogue is used to access novels. 58 users searched for interesting novels to read in a simulated situation where they had only a vague idea of what they would like to read. Data consist of search logs, pre and post search questionnaires and observations. Results show, that investing effort on examining results improves search success, i.e. finding interesting novels, whereas effort in querying has no bearing on it. In designing systems for fiction retrieval, enriching result presentation with detailed book information would benefit users.

Categories and Subject Descriptors

H.3.7. [Digital Libraries]: User Issues

General Terms

Human Factors

Keywords

Fiction Retrieval, Novels, Readers, Public Libraries, Search Tactics, Search Effort, Search Success

1. INTRODUCTION

Reading novels is a popular leisure time interest. Fiction was read at least once a year by 50 % of Americans in 2008 [10] and by 80 % of Finns in 2010 [13]. Public libraries are major channels of getting access to novels [9]. Studies on the outcomes of public libraries show that the major benefit derived from their use is the pleasure of reading fiction [6, 15]. Despite this fact, there has not been much interest in studying and developing systems for fiction retrieval since the 1980s [1]. The effort in developing search systems has been focused on retrieving non-fiction [2, 4].

Traditionally library catalogs have supported accessing novels if the reader knows the name of the author or the title of the novel. It is known that about half of the fiction borrowed is found by browsing, half by known item search [14]. This indicates a need to develop systems supporting other fiction search tactics than known item search. There are signs of enriching public library catalogs to include features supporting fiction retrieval like extended book descriptions or indexing [1, 12]. However, the utility of these tools for accessing novels is not studied. Our aim is to analyze how tools provided by an enriched public library

catalogue are used to access interesting novels to read.

2. RELATED RESEARCH

Next we introduce studies on how readers access fiction in libraries and on evaluation of fiction search systems. The literature in this field is scarce [1]. In [8], Pejtersen summarizes her seminal works in fiction retrieval. As far as we know, there have been no published studies on fiction searching in commercial sites like Amazon. The discussion in [3] hints also to that.

Goodall [5] differentiates two stages in the book search process in the library. Readers identify first attributes in the books, which trigger their interest, and after that focus on attributes, which generate the decision to borrow the book. In the filtering stage, external attributes of the book like its cover or title are perceived as important, whereas in the selection stage, internal attributes of the book like text on the back of the cover or passages of the text in the book are considered as useful. Ross [11] has made a roughly similar distinction based on interviewing 194 committed readers. She distinguished between the clues in the book and elements in the book as indicators of an interesting book.

Pejtersen [8] has defined three major tactics for accessing fiction, which match to our research goals. Analytical search strategy is used when readers wish to find novels about some topic like the Second World War. Search by analogy is generated when readers want something similar to novel X, e.g. a novel they had previously read. Browsing strategy is applied in situations when readers have only a vague idea of what they would like to read. They are simply browsing for finding a good novel.

Based on observing user-librarian negotiations for finding fiction, Pejtersen [8] has designed a fiction search system called the Book House. It consisted of facets representing various attributes of novels as perceived by library users. These facets were access points to novels. The evaluation showed that the system was useful and pleasurable to use [8]. All the available system functionalities were used and the fiction classification system fully accepted. The users found it useful in finding novels.

3. RESEARCH DESIGN

The aim of this study is to analyze how an online catalog in a public library is used for finding novels to read. We focused on a situation when the readers have only a vague idea of what they would like to read. This corresponds to the browsing strategy in Pejtersen [8]. In addition to known item search, browsing is the second major strategy for accessing fiction [8, 14]. Conceptually, browsing includes also similarity search and category search, because in these search modes the reader does not know exactly

what she wants. Browsing may lead to similarity search and category search. Therefore, we chose browsing as the search mode in our study. The specific research questions are:

- What kind of search moves were used for accessing novels?
- Was there an association between moves and search success?

PIKI library system serves several municipalities in Tampere region in Finland. It includes a database containing metadata about the books in the networked libraries, and an interface to interact with that information and search books. The metadata for fiction contains typical bibliographic information added with keywords from the fiction thesaurus “Kaunokki” [12] and tags assigned by users. The metadata includes also images of book covers, recommendations by users and librarians, and availability information. The object of a default search is the whole database. Search results are ranked by relevance, but they can be ordered also alphabetically by author or title, and by publication year. Search results can be limited by category, i.e. fiction vs. non-fiction, by the type of material like book, video, etc., by keyword, by language, or by library. Clicking the book title on the result list reveals the metadata of the book with availability information.

In addition to author, title, free term or keyword search, users may start from recommendation pages. They include various lists of books and recommendations by users and librarians. Users can also search for similar books based on keywords.

For the study 58 participants were recruited in May 2011 from three public libraries of various sizes in PIKI area. Of the study subjects, 26 were recruited in a big main library, 22 in a medium sized main library and 10 in a small branch library. 36 were females and 22 males. Their age varied between 14 and 70 years, the average age being 34 years. They were relatively highly educated, 39 % had a university degree, and 23 % had a high school education, and the rest had a lower education. They read on average 24 novels per year ranging from 0 to 120 novels.

The search task was as follows: You are in a library in a situation when you do not have a clear idea of what you would like to read. Please use the PIKI catalog to search for a novel of interest to you, which you would like to read. Do not search for a particular author or novel, although you may use this as a point of departure for your search. Thus, we simulated a typical browsing situation [5, 11] when readers have only a vague idea of what they would like to read [8]. The search was ended when an interesting novel was found, or when the searcher gave up the search task as unsuccessful.

The search screen was recorded. The researcher observed the search sessions and made notes. The searchers filled in a pre-search questionnaire eliciting demographic information, information about reading orientation, the use of the library and search tactics for books in the library. After the search they filled in a post-search questionnaire including a pattern of questions for assessing various features of PIKI interface, ranking of the novel found and open questions concerning the criteria of selecting the novel and the difficulty of the search task.

Search moves were observed from the recordings of search screen. 29 move types were identified. A move is an identified use of a system feature like a keyword search, an author search, inspecting result list, limiting it, or exploring book metadata. The number of the moves varied from 2 to 21. The distribution of

moves was very scattered. The four most common moves were book clicks (20.4 %), result list (20.2 %), free text search (8.2 %) and category limitation (6.5 %). The proportion of all other 25 moves varied between 4.8 % and 0.2 %. Therefore, for the economy of analysis we collapsed similar move categories like field search (by publication date, library, language, category, material) or limiting result list (by keyword, language, etc). We also recorded the time used for the search.

The indicator of the success of search was an interesting novel found. The searchers rated the novel in a three-point scale from one to three (least to most interesting). If the searcher could not find an interesting novel, the scoring was zero.

4. RESULTS

When starting a search, readers could select either a quick search, an advanced search or a recommendation page as their point of departure. Quick search consists of a search box with a drop down menu suggesting a keyword with information about its type like author when keying in search terms. In an advanced search it is possible to formulate a query by selecting several fields to search. Recommendation pages include various lists of books and recommendations with links.

Advanced search was the most popular search mode (72.4 %) followed by quick search (19 %) and recommendations (17.5 %) (table 1). Readers made on average 7.9 moves when attempting to find a good novel. Of these moves on average 3 were advanced searches, 0.4 quick searches and 0.5 recommendation moves. Users retrieved on average 1.6 result lists, and limited these result lists 0.6 times. On the result lists they clicked 1.6 books, but read only 0.2 book descriptions containing more than bibliographic data. The average interest score of the book accepted was 2.4. The average search time was 215 seconds.

Table 1. Basic statistics of the main study variables (n=58)

Variable	Mean	Stddev	Min	Max	% using
Quick search	0.4	1.1	0	6	19.0
Advanced search	3.0	2.9	0	12	72.4
Result list	1.6	1.4	0	6	86.2
Result list limit.	0.6	1.3	0	6	27.6
Book clicks	1.6	1.3	0	7	95.1
Book description	0.2	0.6	0	3	10.3
Recommendation	0.5	1.3	0	7	17.5
All moves	7.9	4.3	2	21	100
Book scores	2.4	0.9	0	3	100
Search time	215	118	76	593	100

As table 2 indicates, the most popular search tactic was field search (63.8 %) followed by free term search (44.8 %). Known item search and keyword search were equally popular.

An average search was relatively short consisting of about eight moves and lasting about 3.5 minutes. A typical search consisted of advanced searches including mostly field searches or searches with terms from controlled or free text vocabulary. Searchers seldom limited the result list, but immediately assessed novels by

examining bibliographic book information. They explored very seldom more detailed book descriptions for assessing novels' value. The searches can be considered as successful. Only five searchers out of 55 could not find a novel, which they considered as interesting. Evaluation scores in three cases were missing. Thus, 50 searchers had a successful result, i.e. a novel rated at least with value one. Of the searchers only one rated the novel with value one, nineteen with value two, and the rest thirty with value three. Thus, about 55 % of the searchers retrieved a novel with the highest interest rank.

Table 2. Basic statistics of the search tactics variables (n=58)

Search Variable	Mean	Stddev	Min	Max	% using
Known item	0.6	1.1	0	5	32.8
Free term	0.8	1.3	0	6	44.8
Keyword	0.5	1.0	0	5	32.8
Field search	1.4	1.4	0	5	63.8

We were curious to know whether the search process variables were associated to the success of search measured by the interest rate of the novel found. We analyzed the association between search moves and search success by calculating Pearson correlation coefficients. The results indicated that none of the search process variables in tables 1 and 2 excluding the result list was significantly associated with the perceived value of the novel. The number of result lists visited correlated significantly with the success ($r=.28$; $p=.04$). Thus, it seems that search success was not associated with the search moves or their combinations used excluding the number of visits in the result list.

Success was neither associated with search effort measured as time used in searching ($r=-.14$; $p=.31$) or the total number of moves ($r=.23$; $p=.10$). However, we observed that effort invested in exploring the search results and in querying were significantly associated with the search success. Correlation between the time invested on an average move and the interest rating of a novel found was $-.45$ ($p=.001$) (table 3). Thus, quick shifts from move to move predict finding an interesting novel. The correlations show also that the greater proportion of the moves devoted to looking at the result list ($r=.34$; $p=.013$) or examining novels in detail found on the result list ($r=.31$; $p=.022$), the more likely searchers found an interesting novel. Deviating from this finding, the proportion of quick and advanced searches of all moves was negatively associated with the ratings of the novels selected ($r=-.27$; $p=.045$). Thus, the greater the proportion of quick or advanced search moves of all moves, the less interesting novels were found.

In all, these findings hint, that search formulation variables, i.e. querying, were not associated with finding an interesting novel to read, and their great proportion of all moves contributed to an unsuccessful search result. The proportion of moves devoted to exploring result lists and book information, however, helped searchers to find interesting novels. Thus, the more swiftly the searchers proceeded from move to move, but the more effort they invested in exploring results list and book information, and the less effort in search formulation moves, the more interesting novels they found. The findings imply, that search formulations are less important than examination of search results as conditions for finding an interesting novel to read.

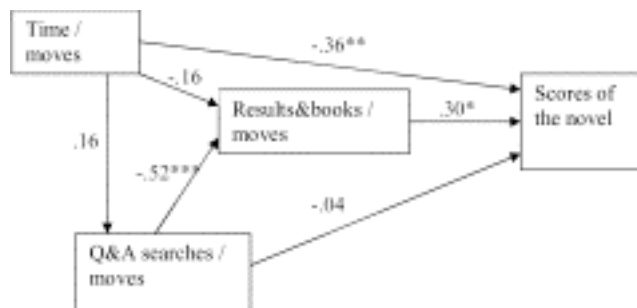
Table 3. Correlations between the average time per move, search effort and the interest grade of a novel (n=58)

Variables	Book scores	Time/moves	Results/moves	Results, book/mo
Time/moves	-.45**			
Results/mo	.34*	-.19		
Results, book/moves	.31*	-.24	.70***	
Q&A searches/mo	-.27*	.16	-.03	-.54***

Legend: * = $p < .05$; ** = $p < .01$; *** = $p < .001$

The previous correlation analyses suggest that the following variables were significantly associated with search success: the average time per move, result lists per move, results and book information per move, quick and advanced searches per move. We use these variables for predicting search success, i.e. the rating of the novel found. Because the two variables measuring the proportion of result list exploration of all moves were conceptually correlated, we removed the variable measuring only visits in result lists, and kept that one which included also exploring book information. The latter one reflects more validly the effort put in exploring the search results.

The model building aims at analyzing the direct and intermediated effects of each independent variable to dependent variable. The model indicates the relative effect of each variable to other variables, i.e. it indicates the effects other variables controlled [7]. Path analysis was used for testing the model. In the path analysis standard regression coefficient are used [7]. The model (figure 1) was significant ($F=7.14$; $p=.000$) indicating a good fit with the data. The multiple correlation (R) of the model was .548, and adjusted R squared .258. Thus, the model explains about 26 % of the variance in the scores of the novels.



Legend: * = $p < .05$; ** = $p < .01$; *** = $p < .001$ (n=58)

Figure 1. A path model for predicting the scores of the novel

The path analysis indicates that time used per move has a significant direct effect on the scores of the novel found ($\beta = -.36$). Also the proportion of search result exploration of all moves has a significant effect on novel scores ($\beta = .30$), whereas the proportion of quick and advanced searches of all moves has no effect on the interest rating of the novel ($\beta = -.04$). The average time per move has a significant effect neither on the proportion of results exploration ($\beta = -.16$) nor on quick and advanced searches of all searches ($\beta = .16$). Interestingly, the proportion of quick and advanced searches has a very large significant effect on the variation in the proportion of result exploration ($\beta = -.52$).

In all, the model indicates, that the less time the searcher used per move, the more interesting novels were found. The average time used per move did not have a significant influence on the proportion of moves devoted either to search formulation or the results exploration. Although these beta coefficients were not significant, their directions hint, that the less time used per move, the more effort was invested in examining the result lists and books information, and the less effort in search formulations. In addition, the more effort put in querying, the less effort allocated in examining results. Thus, it seems that there was a bifurcation of search strategies emphasizing either querying or result list examination. These two strategies had very different effects on finding an interesting novel. Investing effort on examining the result list and book information has a significant positive effect on finding an interesting novel, whereas emphasis on search formulations has no bearing on finding an interesting novel.

5. DISCUSSION AND CONCLUSIONS

As far we know, this is the first study since Pejtersen [8] to analyze the search tactics used by readers for accessing fiction in enriched public library catalogs. We observed how readers searched for an interesting novel in a situation where they had only a vague idea of what they would like to read [8]. We found out that the use of various moves for searching novels was scattered. The most common moves were advanced search, browsing result list and examining book information. The use of various moves was not associated with the success of the search, with finding an interesting novel. However, it turned out that the less time used per move, and the greater the proportion of moves for examining the result list and book information, the more interesting the novel found. The proportion of search formulation moves was not associated to the search success. The model build hints that readers used two alternative strategies with differing success for accessing good novels. The strategy emphasizing search formulations was not associated with finding an interesting novel, whereas the more effort invested in examining results in the search, the more interesting novel was found.

Effort invested in exploring search results instead of querying is an essential factor for finding interesting novels in a situation when readers do not have a clear idea of what they wish to read. Although readers have only a vague idea of the object of interest, they know genres, authors and titles, and have attributes of good novels in their mind [11]. They use this information when selecting books to read. It is likely that what is considered as an interesting novel varies a lot in the sense that the substitutability of novels is great in this situation. Several alternatives may do, not only one. Therefore, effort put on exploring the result list is more productive than querying in the search for good novels to read.

Our results suggest that in designing systems for fiction retrieval, it is important to enrich result list presentation. Readers need more clues about where to infer that the novel could be of interest, and also more options to be informed about the content of the novel [5, 11]. The latter include e.g. recommendations by fellow readers and librarians, texts on the back of the books and links to critics of the novels and to author information like in some electronic bookshops.

It can be supposed that the more readers know about literature, the more effectively they can identify interesting fiction [11]. In the studies to come, we analyze whether readers' literary competence is connected to fiction search process and output. Also

experimental studies on evaluating new tools for supporting fiction retrieval are needed.

6. REFERENCES

- [1] Adkins, D. & Bossaller, J. E. 2007. Fiction access points across computer-mediated book information sources: a comparison of online bookstores, reader advisory databases, and public library catalogs. *Library & Information Science Research* 29(3), 354-368.
- [2] Case, D. 2007. *Looking for information. A survey of research on information needs, seeking and behavior*. 2nd ed. Academic Press.
- [3] Buchanan, G. & McKay, D. 2011. In the bookshop: examining popular search strategies. In *Proceedings of the 11th Joint Conference on Digital Libraries*. New York: ACM Press, 269-278.
- [4] Elsweiler, D., Wilson, M.L. & Lunn, B.K. 2011. Understanding casual leisure information behavior. In: Spink, A & Heinström, J. (Eds). *Future directions in information behaviour*. Emerald, 211-241.
- [5] Goodall, D. 1989. *Browsing in the public libraries*. LISU Occasional paper No 1. Loughborough: Library and Information Statistics Unit.
- [6] Lance, K.C., Steffen, N.O., Logan, R., Rodney, M.J., & Kaller, S. 2001. *Counting on results: new tools for outcome-based evaluation of public libraries*. Aurora, CO: Bibliographical Center for Research. Retrieved in December 20, 2009 from <http://www.lrs.org/documents/cor/CoRFin.Pdf>.
- [7] Pedhazur E. 1982. *Multiple regression in behavioral research*. 2nd ed. New York: Holt, Rinehart & Winston.
- [8] Pejtersen, A.M. 1989. *The Book House. Modeling user's needs and search strategies as a basis for systems design*. Roskilde: Risø National Laboratory.
- [9] *Perceptions of Libraries, 2010: Context and Community*. (2010). OCLC. Retrieved August 10, 2011 from <http://www.oclc.org/reports/2010perceptions.htm>.
- [10] *Reading on the Rise* (2008). National Endowment for the Arts. Retrieved November 27, 2011 from www.nea.gov/research/Readingonrise.pdf
- [11] Ross, C.S. 2001. Making choices: What readers say about choosing books to read for pleasure. *The Acquisition Librarian* 13(25): 5-21.
- [12] Saarti, J. & Hypén, K. 2010. From thesaurus to ontology: the development of the *Kaunokki* Finnish fiction thesaurus. *The Indexer* 28(2): 50-58.
- [13] Serola, S. & Vakkari, P. 2011. *The public library in the activities of people*. In Finnish. Publications of the Ministry of Culture and Education 2011/21. MCE: Helsinki.
- [14] Spiller, D. 1980. The provision of fiction for public libraries. *Journal of Librarianship* 12(4): 238-266.
- [15] Vakkari P. & Serola, S. 2012. Perceived outcomes of public libraries. *Library & Information Science Research* 34(1): 37-44. [doi:10.1016/j.lisr.2011.07.005](https://doi.org/10.1016/j.lisr.2011.07.005)