The role of readers' literary preferences in predicting success in fiction search

Pertti Vakkari
Faculty of Information Technology and Communication Sciences,
Tampere University, Tampere, Finland, and
Anna Mikkonen
Häme University of Applied Sciences, Hämeenlinna, Finland

Abstract

Purpose – The purpose of this paper is to study what extent readers' socio-demographic characteristics, literary preferences and search behavior predict success in fiction search in library catalogs.

Design/methodology/approach – In total, 80 readers searched for interesting novels in four differing search tasks. Their search actions were recorded with a Morae Recorder. Pre- and post-questionnaires elicited information about their background, literary preferences and search experience. Readers' literary preferences were grouped into four orientations by a factor analysis. Linear regression analysis was applied for predicting search success as measured by books' interest scores.

Findings – Most literary orientations contributed to search success, but in differing search tasks. The role of result examination was greater compared to querying in contributing search success almost in each task. The proportion of variance explained in books' interest scores varied between 5 (open-ended browsing) and 50 percent (analogy search).

Research limitations/implications – The distribution of participants was biased toward females, and the results are aggregated within search session, both reducing the variation of the phenomenon observed.

Originality/value – This study is one of the first to explore how readers' literary preferences and searching are associated with finding interesting novels, i.e. search success, in library catalogs. The results expand and support the findings in Mikkonen and Vakkari (2017) concerning associations between reader characteristics and fiction search success.

Keywords User studies, Reading, Books, Modelling, Online catalogues, Interest criteria, Fiction search, Search success

Paper type Research paper

Introduction

An interest in various aspects of fiction book search has gained some footing among studies on information retrieval. This change reflects the calls for broadening the scope of the research field to cover also casual leisure information searching in addition to factual, work-related searching (Elsweiler *et al.*, 2011). The object of retrieval broadens from information objects with factual content to information objects with fictitious content. The nature of activity producing information needs expands from instrumental to expressive, too. An instrumental activity is a means to a goal, whereas an expressive activity is a goal as such, i.e. an activity is valuable for its own sake (O'Connor, 1987).

Reading fiction is often considered as an end itself, pleasurable as such (Ross, 2001). The differences mentioned above are reflected in the search behavior, in the nature of criteria for selecting interesting books and consequently, in search tactics (Mikkonen and Vakkari, 2017;

Pejtersen, 1989; Ross, 2001; Vakkari and Pöntinen, 2015). While topicality is considered typically the major criterion for retrieval effectiveness (Schamber, 1994), it is a poor indicator of success in fiction search compared to other characteristics of novels like genre, plot or literary style (Mikkonen and Vakkari, 2016b; Pejtersen, 1989; Ross, 2001; Vakkari and Pöntinen, 2015). The major search scenarios differ also between fiction and non-fiction searching (Mikkonen and Vakkari, 2016a; Pejtersen, 1989; Ross, 2001).

The studies on fiction search have dealt with visualization techniques for book search (Thudt *et al.*, 2012), readers' means of accessing fiction in the public library (Mikkonen and Vakkari 2012; Saarinen and Vakkari, 2013) or in bookshops (Buchanan and McKay, 2011), selecting fiction in library catalogs and other sources (Mikkonen and Vakkari, 2016a; Pejtersen, 1989; Vakkari and Pöntinen, 2015; Tang *et al.*, 2014), relevance and interest criteria in selecting fiction (Koolen *et al.*, 2015; Mikkonen and Vakkari, 2016b) and reader characteristics and fiction searching (Mikkonen and Vakkari, 2017).

Studies on fiction book search do not typically differentiate between readers' literary preferences as factors, which may influence search behavior (Mikkonen and Vakkari, 2017). As topicality is not a valid indicator for search success (Mikkonen and Vakkari, 2016b; Pejtersen, 1989; Vakkari and Pöntinen, 2015), it is essential to find out which characteristics of literature readers prefer when looking for and selecting interesting novels. There are only a few studies exploring readers' literary preferences and search behavior in electronic (Mikkonen and Vakkari 2017) and traditional (Ross, 2001; Saarinen and Vakkari, 2013; Spiller, 1980) library environment. There is a lack of studies about how literary preferences are associated with fiction search behavior. It is likely, however, that readers with differing interest profiles differ also in their search tactics, interest criteria for selecting novels (Mikkonen and Vakkari, 2017; Pejtersen, 1989; Saarinen and Vakkari, 2013), and consequently search success. If we can better reveal readers' interest profiles and their associations to search behavior, we might be able to better serve their attempts to find good reads.

We have studied earlier various aspects of fiction book search in library catalogs including how readers' literary preferences relate to the patterns of fiction search. Literary preferences were divided into two categories and search behavior was analyzed variable by variable in these two categories (Mikkonen and Vakkari, 2017). The study at hand uses the same data set. It enriches our earlier findings by clustering readers' literary preferences into four categories and by showing by linear regression analysis to what extent each literary preference and other factors in the models predict separate and jointly search success. Our contribution consists of providing a richer categorization of literary preferences and multivariate models including literary preferences and other factors predicting fiction search success. Information about the differences in readers' preferences and consequent differences in search behavior could be used in developing tools in library catalogs to better support varying reader groups to find novels to read.

The aim of this study is to explore how readers' literary preferences are associated with fiction search behavior in library catalogs. In particular, the aim is to analyze to what extent readers' socio-demographic background, literary preferences and search behavior predict the success of finding interesting novels for various search scenarios in library catalogs.

Related research
Two types of studies are relevant in this context. Studies on reader characteristics, literary preferences, in particular shed light on features, which may be associated to fiction search. Studies on searching novels in library catalogs and libraries provide results on fiction search and selection patterns for backing this study.

Reader characteristics
Characteristics that affect the tendency to read fiction have been identified in several
studies. It has been shown that readers' gender, age and educational levels are associated
with the tendency of reading fiction books. Women read fiction more frequently and also
more varied types of books than men (Ross, 2001; Stockmans, 2003). Studies also show that
reading fiction decreases by age (Ross *et al.*, 2006), while it increases by the increasing level
of education (Kraaykamp and Dijkstra, 1999; Ross *et al.*, 2006). An increase in educational
level is typically associated with the type of book genres read (Kraaykamp and Dijkstra,
1999; Ross *et al.*, 2006). A Dutch survey showed that the complexity and prestige of books
read increased as readers' educational level increased (Kraaykamp and Dijkstra, 1999).
Complexity meant that books were demanding to read and that reading required literary
prior knowledge. Prestige referred to canonized books with high literary value. Complexity
and prestige were assessed by literary experts. The authors concluded that education both
socializes to and provides means in reading complex literary works.

Some studies touch upon the readers' varying reading preferences, i.e. what is expected
from reading (Miesen, 2003). These include fulfillment of affective needs like enjoyment and
entertainment as well as utilitarian outcomes such as reading fiction for learning and
practical knowledge (Miesen, 2003; Ross, 2001; Usherwood and Toyne, 2002). Miesen (2003)
studied adults' reading motives with a questionnaire measuring behavioral beliefs
toward literary reading. The study revealed five major factors of motives for fiction reading:
affect – enjoyment, utility – intellectual development, utility – broadening one's horizon,
prestige – self-cohesion and relief from boredom.

Saarinen and Vakkari (2013) categorized readers into three types based on their
reading motives and types of novels read. Escapists were seeking relaxation and
distraction from daily routines from pleasure reading. They identified themselves with the
characters and the plot of the novel. They typically read literature belonging to genres like
thrillers or romances. Esthetes read high standard novels observing language use and
aspects of narration with the aim of developing themselves and receiving new experiences
and views. Realists expected from novels realism and credible descriptions of everyday
life or development of a subject. Their motivation to read was learning new things and
also relaxation with lighter reading than non-fiction. When browsing for finding good
books escapists leaned on library's genre classification and the shelves of returned novels,
while esthetes and realists browsed shelves of new books in addition to shelves of
returned books.

Searching fiction in library catalogs
Searching fiction. When readers looked for good novels without clear goals in an online
catalogue, effort invested in examining result lists and book metadata was positively
associated with finding interesting novels, whereas effort in querying had no bearing on it
(Oksanen and Vakkari, 2012).

Mikkonen and Vakkari (2016a) found out that users needed more queries, more search
moves and opened more book pages to find equally interesting books in a traditional library
catalog compared to an enriched library catalogue.

Based on a pattern of questions concerning reading preferences of fiction, Mikkonen and
Vakkari (2017) clustered readers into aesthetes and entertainers. The former preferred artistic
and esthetic pleasures of reading, whereas the latter enjoyed escape and comfort provided by
reading. The search patterns of these two groups differed somewhat in various search
scenarios. In known author search and in open-ended browsing for good novels, esthetes
devoted time for browsing SERPs and then clicked a few book pages to find interesting
novels, whereas entertained readers clicked several items on SERPs to find a good book.

The authors concluded that esthetes were able to infer from the result list, which books could be of interest, whereas entertained readers made the decision more by trial and error by clicking several items. The authors suggest that this difference is due to readers varying literary knowledge. In the search by analogy (i.e. similarly interesting novels), there were no differences in search patterns between the groups. Both groups made several SERP visits and opened many book pages for finding interesting books.

Selecting fiction. It is evident that selecting fiction differs from selecting non-fiction. Studies on relevance show, that topicality, what the document is about, is the major criterion in selecting non-fiction (Schamber, 1994). The topic of a novel plays a minor part in selecting fiction. Readers typically select novels of the authors they know. If they have no author in mind, they focus on the genre, plot, setting, characters or literary style of a novel (Mikkonen and Vakkari, 2016b; Pejtersen, 1989; Ross, 2001; Saarinen and Vakkari, 2013). The substitutability of books differs considerably between fiction and non-fiction, too. Topicality restricts the section of non-fiction to certain items, whereas a novel can be of any theme if the major criterion like genre or literary style is appropriate.

By interviewing 500 library users in the UK, Spiller (1980) found that 54 percent searched for novels by title or author's name, and the rest 46 percent by browsing in the library. Almost 78 percent of fiction searches were carried out by combining known author or known title search with browsing.

Based on a survey of the adult population in Finland, Mikkonen and Vakkari (2012) analyzed readers' methods of accessing fiction books in public libraries. The most common method was known book or author search, which was used often by 57 percent of the respondents. It was followed by browsing the shelves (29 percent) and skimming the returned loans (27 percent).

Pöntinen and Vakkari (2013) studied how readers explore metadata in book pages when selecting fiction in two library catalogues. They analyzed eye-movements of 30 participants selecting fiction for four search tasks. They found out, that although participants devoted most attention in book pages to content description and keywords, these had no bearing on selecting an interesting novel. Author and title information received less attention but were significant predictors of selection.

In all, readers search and select fiction in libraries either by searching known titles or known authors, or by browsing, when they do not have a particular author or title in mind (Mikkonen and Vakkari, 2012; Pejtersen, 1989; Spiller, 1980). Most searches are based on known author or title, while browsing for interesting novels occurs more rarely (Mikkonen and Vakkari, 2012; Ross, 2001; Spiller, 1980). When selecting fiction in catalogs, readers devote more attention to SERP and book pages instead of querying for finding interesting books (Oksanen and Vakkari, 2012). Readers' increasing educational level enhances their literary knowledge, which seems to be associated with their ability to identify interesting authors and titles, and thus, search tactics (Kraaykamp and Dijkstra, 1999; Mikkonen and Vakkari, 2017; Ross, 2001; Saarinen and Vakkari, 2013).

Research design
The study seeks to answer to the question to what extent readers' socio-demographic background, literary preferences, search actions, the characteristics of search tasks and catalog type predict search success as measured by books' interest scores.

Participants
In total, 80 participants with fiction reading interest were recruited in public libraries and reading circles. The snowball method and newspaper advertisement were used. Participants were randomized into two groups. Half of the participants completed five search tasks in a

typical public library catalog "Sata," while the other half used "Sampo," a catalog designed for fiction search. In total, 18 percent of the participants were male, and 82 percent female. The age distribution varied from 20 to 80 years. 18 of the participants had a middle level education, while 82 percent has a high-level education. In order to increase the variance of educational level, it was divided into four classes in this study.

Search tasks
Four search tasks were designed simulating typical situations in selecting fiction in public libraries (Goodall, 1989; Pejtersen, 1989; Ross, 2001; Spiller, 1980). For the search tasks, the participants could use only the assigned catalog for finding books:

- Known author search: "A friend recommends you to familiarize yourself with the novels of Olli Jalonen. Find Olli Jalonen's novels and choose two novels which are of interest to you."

- Topical search: "Find three novels of interest about upper class life in the 19th century."

- Open-ended browsing: "Find three novels that interest you which you would like to read."

- Search by analogy: "Mention one novel that you have read and found interesting recently. Now search for three novels that you would consider similarly interesting as the one you mentioned."

The first task was author search, which is the most common method of searching for novels in public libraries (Goodall, 1989; Mikkonen and Vakkari, 2012; Spiller, 1980). Hints from friends and relatives are important sources of book information (Ross, 2001). We used this idea for framing this task. The second task was a topical search, where the reader wishes to find a novel on a particular theme. This method was identified in Pejtersen (1989). The aim of the third task was to generate browsing, which is a popular method of choosing fiction (Goodall, 1989; Mikkonen and Vakkari, 2012; Spiller, 1980). The fourth task simulates a situation when a reader seeks to find a novel, which resembles an appealing one read earlier. This kind search by analogy is common among readers (Ross *et al.*, 2006).

Catalogs
Kirjasampo (Sampo) is an enriched online service for fiction literature, and Satakirjastot (Sata) is a traditional online public library system. Both are real systems in use.

Sampo is a fiction literature portal based on principles of the semantic web. In addition to bibliographic information about the books, also content and context information is indexed in the database. It employs functional content-centered indexing, ontological vocabularies and the networked data model of linked data (Hypén and Mäkelä, 2011). The Finnish fiction ontology Kaunokki is used for indexing the works in the database (Saarti and Hypén, 2010).

Sampo has two basic functionalities, searching and browsing. Searching includes the possibility of using text or cover image queries. Users may also utilize book recommendations at the main page or select books through fellow users' bookshelves. As a result of text querying, a list of categories "author, person or other actor," and various "genres" like novels, short stories or cartoons are provided. After having clicked a category, the user is provided with brief information about the book including the title, cover image and a few keywords. By clicking a book title the user is transferred to the information page of the book, which includes the following metadata: author, title, keywords from facets, genre, theme, figures, time and place and other keywords, a content description of the story

(typically from the back of the book), a sample text passage, publication data, cover image, possible presentation by other readers and "see also" recommendations.

The browsing interface provides the user with a possibility to wander through the context of a work and through it to other works. Besides allowing one to walk the semantic network through the actors, books and keywords, the interface also provides recommendations, which automatically locate interesting semantically related content related to the viewed work (Hypén and Mäkelä, 2011).

Sata is a traditional public library online catalog providing users with quick search, advanced search and a browsing option. Searching starts with querying. In quick search users key in search terms in a textbox, whereas in advanced search in addition to that they may limit the search by the type of literature (fiction – non-fiction), author, title, keywords or other bibliographic information.

The result list is organized by material type like books or CDs. In the list, the user should click the link "book" in order to explore the list of books retrieved. The list includes the following metadata from each book: author, title, material type, publication year and library class. A click on a book title leads to the book page containing metadata title, author, publication data and keywords from the fiction thesaurus Kaunokki. The upper right corner includes more recent books, a cover image and a link to the content description of the story.

In Sampo, users may end up on a book page without querying, e.g. by browsing via links, whereas in Sata one query is a minimum requirement for accessing a book page. In general, book pages in Sampo include more information about the books compared to Sata. The characteristics of both catalogs are presented in Table I.

Experimental procedure

The experimental setting was pre-tested with one participant to gain information on the duration of the test as a whole and to see if the instructions were unambiguous enough.

The search logs were saved with a Morae Recorder. It captures audio, video and on-screen activity during a research session. Variables measuring the search actions were manually calculated from on-screen activity after conducting the user tests. The audio was used in analyzing and classifying participants' search queries.

In the tasks, the participants were asked to search for novels that were of interest to them. "Known author task" functioned as a training task at the beginning of the experiment. The time for completing the tasks was not limited. Latin square rotation was used with the tasks. During the experiment, the researcher was present to help in case technical problems occurred.

Characteristics	Sampo	Sata
Searching		
Quick search	Yes	Yes
Advanced search	No	Yes
Image search	Yes	No
Browsing	Yes	Yes, after querying
Book page in-formation		
Author, title	Yes, always	Yes, always
Keywords	Yes, often	Yes, often
Content description	Yes, often	Yes, seldom
Publication data	Yes, always	Yes, always
Cover image	Yes, often	Yes, seldom
Other	Bookshelves, tip of the day, literary news, book reviews, recommendations	Library class

Table I.
The major characteristics of catalogs Sampo and Sata

Based on Miesen (2003) and Saarinen and Vakkari (2013), a pre-questionnaire was designed, which measured participants' socio-demographic characteristics and literary preferences. Literary preference refers to the characteristics of novels, which are appealing to the readers (Miesen, 2003). In the post-task questionnaire, participants were asked to rank the novels found according to how much they were of interest to them. They also rated on a four-point scale to what extent they know the production of the author in author search (very well – not at all), and how difficult they perceived each search task (very difficult – very easy).

Measures

When selecting novels, readers refer to characteristics other than topicality like genre, plot or literary style, which are appealing or of interest to them (Mikkonen and Vakkari, 2016b; Pejtersen, 1989; Ross, 2001; Saarinen and Vakkari, 2013). Therefore, the success of search tasks was measured by the books' interest grading by the participants. They were asked to indicate on a three-point scale how interesting they perceived the novel. The grading was "very interesting" (3), "somewhat interesting" (2) and "little interesting" (1). If the no interesting novel was found, the scoring was 0. The sum of interest scores for each task were calculated. Although the measurement level of this variable is ordinal, we treat it more like an interval level variable. This decision makes the analyses more economic by reducing the number of operations needed for producing the results. However, we do not assume that the intervals within the variable are equal. The study variables are listed in the following list:

(1) Socio-demographic factors:
- Gender, Age, Educational Level.

(2) Book preferences:
- four literary orientations; and
- the number of fiction books read per year.

(3) Search actions:
- the number of search moves;
- the number of queries;
- the number of SERP visits;
- the number of opened book pages;
- the number of pivot browsing actions;
- query time;
- dwell time on SERP; and
- dwell time on book pages.

(4) The characteristics of search tasks:
- the perceived difficulty of a search task;
- familiarity with author's production (Author search); and
- topic familiarity (Topical search).

(5) Catalog type:
- Sata/Sampo.

(6) Search success:

- The sum of book's interest scores per task.

A search move is a basic action that advances search process (Bates, 1979). Pivot browsing consists of search moves which re-orientate browsing to follow features such as virtual bookshelves, a tag cloud, a cover image carousel and recommendations.

Our earlier studies (Mikkonen and Vakkari, 2017) showed that readers' search success in both catalogues did not differ. Due to the limited number of participants, data from both catalogs were merged when the regression analysis indicated that the catalog type did not have an effect on search success. The number of cases varied somewhat in the analyses either due to missing information in certain variables or due to removed outliers.

Reader groups

Participants' literary preferences were measured by a pattern of questions, which was based on Miesen (2003) and Saarinen and Vakkari (2013). The questions measured by a four-point scale how important (very important – not at all important) participants considered 12 characteristics of novels. A factor analysis was applied for distinguishing reader groups with differing literary preferences. The number of factors was selected based on their theoretical significance and the clarity of conceptual meaning (Hair *et al.*, 2010). A principal component analysis with varimax rotation was used. Factors with an eigenvalue of at least one were extracted.

The four factors solution explains 70.8 percent of the total variance of constituent variables. The first factor covers 29.2 percent of the total variance. The highest loadings of characteristics refer to a preference to classical novels with history orientation by authors like Bronte or Tolstoy (Table II). They typically include a surprising plot, rich characterization of protagonists, detailed presentation of novel's setting, which is combined with original writing style that may challenge the readers. The first factor reflects the preference to classic novels. It is called classic orientation.

The second factor explains 16.0 percent of the total variation of variables. The highest loadings refer to novels, which influence strongly both readers' feelings and thoughts by skillful language use. Note also that novels that challenge readers have medium high loading and entertaining novels have a negative loading in this factor. This factor reflects artistic enjoyment, and it can be called esthetic orientation.

The third factor covers 14.4 percent of total variance of variables. The novels that are based on real events and represent reality truthfully load highest in this factor. The factor reflects preference to realistic novels. It can be called realism orientation.

Literary characteristics	Factor 1	Factor 2	Factor 3	Factor 4
Based on real events	0.002	−0.080	*0.902*	0.004
Entertaining	0.091	−0.321	−0.235	*0.816*
Skillful and rich language	0.176	*0.801*	−0.126	−0.240
Gripping plot	0.080	0.165	0.225	*0.854*
Arouses feelings	0.016	*0.711*	0.124	0.443
Surprising plot	*0.817*	−0.031	0.103	−0.125
Rich characters	*0.601*	0.302	−0.220	0.153
Thought provoking	0.294	*0.821*	0.049	−0.043
Setting precisely presented	*0.712*	0.000	0.202	0.223
Original style	*0.654*	0.291	−0.115	0.112
Challenges the reader	*0.651*	0.461	0.058	−0.105
Represent facts	0.062	0.092	*0.882*	0.047

Note: $n = 75$

Table II.
The loadings of factors for literary preferences

The fourth factor explains 11.2 percent of the variation of variables. The highest loadings refer to novels, which are entertaining and have a gripping plot. Also novels, which arose feelings have a medium high loading. This factor reflects preference to novels with immersion power. It is called immersion orientation.

The analysis revealed four literary preferences, which were classic, esthetic, realism and immersion orientation toward novels. For elaborating the results, factor scores of these four factors were calculated for each respondent. The factor scores indicate to which extent the factors represent respondents (Hair *et al.*, 2010).

Predicting search success
We analyze next which predictors are associated with books' interest scores in each search task. We used linear regression analysis with enter method for selecting significant predictors for the models. In the enter method, all predictors were first included in the model, and non-significant ones were removed based on researcher's reasoning to build conceptually meaningful models (Hair *et al.*, 2010). Linear regression analysis reveals associations only between independent and dependent variables. Therefore, correlation analysis was applied for analyzing the associations between independent variables.

Author search
The model for book scores in author search is significant ($R = 0.63$; $R^2 = 0.39$; adj. $R^2 = 0.35$; $F = 8.8$; $p = 0.000$) consisting of five variables (Table III). The model explains 35 percent of the variation in the book scores. The model is a combination of readers' search actions, esthetic orientation and familiarity with the author's production.

The stronger the readers' esthetic orientation, the more they know about the author's production, the less search moves they make, the more SERP they visit and the shorter the dwell time in opened book pages, the more interesting books they found.

We can hypothesize that readers with strong esthetic orientation possess literary knowledge, which implies that they also are to a certain extent familiar with the novels by the author, which was the target of the search (Kraaykamp and Dijkstra, 1999; Mikkonen and Vakkari, 2017). These qualities help readers to identify interesting items among the authors' novels. This ability in its turn decreases the number of search moves, which, however, focus on frequent visits on SERP, but which require short dwell times on opened book pages for identifying an interesting novel. It seems that esthetic orientation with knowledge of authors leads readers to browse SERP for identifying a few potentially interesting titles, which they click and select an interesting one by a short visit to the book page.

We validated the previous hypothesizing by checking the data. Esthetic orientation was positively associated with readers' knowledge of author's production ($r = 0.23*$), which is negatively correlated with the number of search moves ($r = -0.32**$), with the number of visited SERP ($r = -0.31**$) and dwell time on book pages ($r = -0.23*$). It seems that esthetic orientation increased the knowledge about authors, which in its turn decreased readers'

Predictors	β
Search moves	−0.50**
SERP visits	0.37*
Dwell time on book pages	−0.21*
Esthetic orientation	0.24*
Knowledge of author's production	0.28**
Notes: $n = 75$. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.10$	

Table III.
A model for
book scores in
author search

search effort for recognizing interesting items among a known author's production. In addition, books' interest scores correlated positively with both esthetic orientation ($r = 0.34$**) and knowledge of author's production ($r = 0.41$***). Thus, esthetically oriented readers are familiar with authors' production, and they tend to find good reads likely regardless of the system they use.

Topical search
The model is significant ($R = 0.47$; $R^2 = 0.22$; adj. $R^2 = 0.17$; $F = 4.9$; $p = 0.002$). It covers 17 percent of the variation in books scores. The model consists of four variables representing search actions, literary preferences, familiarity with the topic and the perceived difficulty of search task (Table IV). The stronger readers' classic orientation toward novels, and the greater their familiarity with the topic, and the easier the search task was perceived, and the longer time spent on SERP, the more interesting novels were found. Interestingly, only one of the four predictors represents search actions, while the rest three represent either the characteristics of the searcher or search task.

In this task, we may also conjecture that readers' literary orientation – classic one – increases familiarity with the search topic, "upper class life in 19th century." Readers interested in classic novels are likely familiar with several established authors dealing with upper class life in nineteenth century such as Tolstoy, James or Eliot. Familiarity with the authors of the topic evidently helps readers to recall and identify pertinent authors and thus makes the task easier. Logically, this all should decrease the dwell time on SERP for identifying matching authors.

A check in data showed that orientation toward classic novels was associated neither with topical knowledge ($r = 0.09$; ns), nor perceived task difficulty ($r = -0.02$; ns). However, task difficulty decreased significantly by increasing topical knowledge ($r = -0.36$***). An increase in topical knowledge decreased somewhat dwell time on SERP ($r = -0.16$; ns), while a decrease in task difficulty significantly decreased dwell time on SERP ($r = -0.39$***). Familiarity with the topic was significantly associated with books' interest scores ($r = 0.26$*), while task difficulty was notably ($r = 0.21$****) associated with those scores. Thus, classic literary orientation did neither increase readers' familiarity with novels about upper class life in nineteenth century, nor the difficulty of searching for such novels. However, the less difficult the task was perceived, the less time was spent on SERP. In addition, increasing familiarity with the search topic and decreasing task difficulty produced higher interest scores for the books. Task difficulty mediated the effect of topical knowledge in finding interesting novels: partial correlation between topical knowledge and books' interest scores controlling for task difficulty was non-significant ($r = 0.19$) compared to the significant original one ($r = 0.26$*).

Open-ended browsing
The model for open-ended browsing included only one predictor, the perceived difficulty of search task. However, the model was significant ($R = 0.25$; $R^2 = 0.06$; adj. $R^2 = 0.05$; $F = 4.5$; $p = 0.037$) explaining only 5 percent of the variation in book scores. The decreasing difficulty of search task increased the chance of finding interesting novels ($\beta = -0.25$*).

Predictors	β
Dwell time on SERP	0.24*
Classic orientation	0.23*
Topical knowledge	0.28*
Task difficulty	−0.26*
Notes: $n = 75$. *$p < 0.05$	

Table IV.
A model for book scores in topical search

One could suspect that open-ended browsing for interesting novels without a clear goal would be more difficult compared to other search tasks, and therefore produce highly scattered searches not associated with interest scores. Empirical evidence, however, did not support this conjecture. The participants did not perceive open-ended browsing more difficult compared to other search tasks as shown by t-tests between those variables. Anyway, readers' search behavior in open-ended browsing for interesting novels was so scattered that there were no patterns linking search actions and books' interest scores.

Analogy search
Patterns in analogy search differed between the two catalogs observed. Results concerning the traditional catalog Sata are presented first.

The model for success in analogy search in Sata was significant ($R = 0.48$; $R^2 = 0.23$; adj. $R^2 = 0.19$; $F = 5.1$; $p = 0.012$) consisting of two predictors (Table V) and covering 19 percent of the variation in books' interest scores. Increasing immersion orientation significantly decreased the success in finding interesting novels, while time spent in query formulation somewhat increased search success. Thus, those who favored immersion had difficulties in finding interesting novels that resemble those, which they have read earlier. However, effort invested in querying somewhat increased the change of finding interesting novels.

The model for books' interest scores in analogy search in Sampo was significant ($R = 0.74$; $R^2 = 0.55$; adj. $R^2 = 0.50$; $F = 10.2$; $p = 0.000$). It explains 50 percent of the variance in book scores. The model consisted of four predictors representing search actions, readers' education and immersion orientation (Table VI). The lower readers' educational level, and the smaller their immersion orientation, the less pivot browsing actions, but the longer dwell time on SERP, the higher the books' interest scores.

It seems that frequent pivot browsing actions (virtual bookshelves, a tag cloud, a cover image carousel) impaired the changes of finding interesting novels that resemble those read earlier, while increasing dwell time on SERP enhanced the changes of identifying this kind of novels. Also decreasing immersion orientation increased the changes to find interesting novels that match with those one has read earlier.

Inter-correlations between independent variables were not significant. However, decreasing immersion orientation correlated somewhat with increasing dwell time on SERP ($r = -0.24$; $p = 0.14$). This hints that those not in favor with immersive novels spent more time in SERP, which increased the changes of identifying interesting novels.

Predictors	β
Query time	0.28**
Immersion orientation	−0.34*
Notes: $n = 37$. *$p < 0.05$; **$p < 0.01$	

Table V.
A model for book scores in analogy search in Sata

Predictors	β
Pivot browsing actions	−0.35**
Dwell time on SERP	0.37**
Education	−0.29*
Immersion orientation	−0.45**
Note: $n = 38$. *$p < 0.05$; **$p < 0.01$	

Table VI.
A model for book scores in analogy search in Sampo

A summary of models

The results show that the pattern of predictors varied greatly by search tasks (Table VII) as did the proportion of explained variance in book scores. Each literary orientation except realism predicted book scores in different search tasks. Esthetic orientation had a positive contribution in author search, and classic orientation in topical search, while immersion orientation had a negative contribution in analogy search. Any orientation did not predict success in open-ended browsing.

Also search actions had a varying effect on book scores across search tasks. Only activities on SERP contributed positively to search success in all tasks except open-ended browsing, while other search actions contributed only in one task each.

Unexpectedly, neither the frequency of reading novels, nor socio-demographic factors did predict book scores. However, educational level was negatively associated with book scores in analogy search in the catalog Sampo. Task difficulty decreased the chances of finding interesting novels, while familiarity with an author's production or with the topic of search, i.e. literary knowledge, increased changes of finding interesting novels.

Discussion

This study is one of the first to explore how readers' literary preferences and searching are associated with finding interesting novels, i.e. search success, in library catalogs. Our results expand and support the findings in Mikkonen and Vakkari (2017) concerning associations between reader characteristics and fiction search success. While Mikkonen and Vakkari (2017) distinguished between esthetes and entertained fiction readers, we categorized readers into four literary orientations: readers with classic orientation, with esthetic orientation, with realism orientation and with immersion orientation. In addition, Mikkonen and Vakkari (2017) focused on search patterns in these two reader groups using univariate, variable by variable analysis, while the study at hand modeled search success by linear regression analysis by including the four literary orientations among predictors in the models. Thus, our results show the role of each literary orientation in predicting search success.

Our results show that all orientations except realism contributed to search success, but each in different search task. The differentiation is plausible, because the search tasks varied a lot from author search to open-ended browsing. Each orientation had a positive effect on book scores in one particular search task except immersion orientation, which had a negative contribution to book scores in one task. Esthetic orientation had a positive effect on books' interest scores in author search as in Mikkonen and Vakkari (2017) and classic orientation in topical search, while immersion orientation contributed negatively to

| | | | | Analogy search | |
Predictor type	Author search	Topical search	Open-ended browsing	Sata	Sampo
Search actions	− Moves, + SERP − Time on book pages	+ Time on SERP		+ Query time	− Pivot br. + Time on SERP
Literary preferences	+Esthetic	+ Classic		−Immersion	− Immersion
Difficulty		−	−		
Familiarity	+	+			
Socio-demogr					− Education
Adj. R^2	0.35	0.17	0.05	0.19	0.50
Notes: −, a negative association; +, a positive association					

Table VII.
The summary of models by search tasks

book scores in analogy search. Open-ended browsing was predicted only by task difficulty. It is likely that browsing for interesting novels without a clear goal in mind was so challenging that search behavior scattered so heavily that there were no associations between search actions and books' interest scores. This may be due to the fact that open-ended browsing occurs rather among bookshelves than in library catalogs (Goodall, 1989; Spiller, 1980).

The mechanisms, which mediated each literary orientation to the search processes, were in most tasks unclear. Only in author search esthetic orientation increased readers' familiarity with the author that was the object of search and helped them to identify interesting titles by that author. This familiarity also reduced the effort in search actions to find good reads (cf. Mikkonen and Vakkari, 2017). In other search tasks, it was not possible to find mediating connections with a literary orientation and search activities. It seems that the literary orientation had a direct effect on book scores. In studies to come it is important to seek to reveal, which factors mediate readers' literary orientations to search actions, and in that way to search success. By knowing these mechanisms, it is possible to inform system design for developing tools to support differing literary preferences in various search tasks. The role of readers' literary knowledge and perceived task difficulty seems to be factors to start with as our results hint.

Excluding open-ended browsing, in almost all other search tasks an increasing effort in examining SERP contributed to books' interest scores, while querying had an effect on book scores only in one task. This finding confirms the finding in Oksanen and Vakkari (2012) and in Pöntinen and Vakkari (2013) that emphasis on examining search results contributes to finding good novels, while querying has a minor role. In addition to search process variables, book scores were also influenced by perceived task difficulty in topical search and open-ended browsing, and topical knowledge in author search and topical search.

The results showed that the perceived difficulty of search task decreased the chances of finding interesting books in topical search and in open-ended browsing. In traditional information retrieval, the perceived difficulty of search tasks also decreases the chances of finding relevant documents (Liu *et al.*, 2012). In topical search, an increase in task difficulty increased dwell time on SERP but decreased the number of book pages visited. Task difficulty increased browsing and scanning of SERP but decreased the number of interesting book pages to click. This finding corresponds to the findings in studies on non-fiction search (Arguello, 2014; Liu *et al.*, 2012).

Familiarity with the author' production in author search and topical knowledge in topical search produced higher book scores and reduced search effort. This indicates that at least in those types of fiction search, which resemble non-fiction search, topical knowledge in a broad sense (literary knowledge in this case) improved search success and influenced search behavior. It has been shown that subject knowledge contributes in several ways to search behavior (Vakkari, 2002; Wildemuth, 2004). In our data, topical knowledge also decreased perceived task difficulty and thus lead to higher book scores in topical search.

Readers' gender, age, educational level or the number of novels read per year were not associated with books' interest scores with one exception. In analogy search, an increase in educational level decreased book scores among users of Sampo catalog. The major finding, however, was that these factors did not seem to contribute to success in fiction search, while readers' literary orientation had an effect on search success, although it varied greatly by task and orientation type. This is somewhat surprising because studies show that gender, education and book reading activity are associated with literary preferences (Kraaykamp and Dijkstra, 1999; Stockmans, 2003). The result also differs from the findings that, e.g., gender has an effect on search (Roy and Chi, 2003) and search result evaluation (Lorigo *et al.*, 2006) patterns on the web. It is evident that more research is needed to test whether our finding is valid.

Our results are limited in a few ways. First, the distribution of gender was biased heavily toward females. This reduces the variance of this variable, and thus, the strength of potential associations with other variables in the study. Second, the results are aggregated within search session, which reduces the variation of the phenomenon observed. Third, open-ended search task, i.e. browsing for interesting novels without a clear goal, may be somewhat artificial. The typical means of accessing novels in libraries are either known author or known title search. Open-ended browsing occurs rather among bookshelves, than in catalogs (Goodall, 1989; Pejtersen, 1989; Spiller, 1980). However, the results in Gäde and Petras (2016) hint that compared to searching, browsing for an open-ended book search was clearly more common both in an experimental book catalog and in a book shop. Fourth, clicking an item in SERP opened a book page, which contained extended bibliographic information, not a full text, which could have been more informative in deciding how interesting the novel was.

Literary orientations were associated with search success in three search scenarios out of four. Supporting various reader groups to find good novels to read implies that fiction search systems should recognize signals corresponding to different literary orientations. The search patterns in the models of this study were either highly scattered or too scarce, with mostly low explanatory power. They do not provide means of inferring literary orientations from search activities and thus, personalize search by the identified orientation. An option to support readers with differing orientations is to ask them to write brief descriptions of the characteristics of novels, which they like and give some exemplary titles. The system would produce a list of novels that match with this description, applying the tools of artificial intelligence. This suggestion corresponds to narrative-driven recommendation, where searchers describe the characteristics of the object of interest and provide examples of it, which a recommender system takes as inputs to generate recommendations (Bogers and Koolen, 2017; Eberhard et al., 2019). Alternatively, library systems could provide readers with pre-defined literary orientations, among which to choose. Each orientation would provide readers with a list of novels matching this orientation. The list could be refined by various additional metadata elements. These solutions imply that novels are indexed so that they can be matched with the pre-defined preferences. It would be of help if a controlled fiction vocabulary is used in indexing like in the systems studied (Hypén and Mäkelä, 2011).

Conclusions

We identified four literary orientations, which were associated with search success, although the pattern of associations varied greatly by orientation and type search task. Surprisingly, neither reading activity nor socio-demographic factors contributed search success. More research is needed for finding mechanisms, which mediate between literary orientations and search actions and success. This information would help designing systems that support readers with differing literary orientations to find interesting novels to read.

References

Arguello, J. (2014), "Predicting search task difficulty", in de Rijke, M. *et al.* (Eds), *Advances in Information Retrieval, Lecture Notes in Computer Science*, Vol. 8416, Springer, Cham, pp. 88-99.

Bates, M.J. (1979), "Information search tactics", *Journal of the Association for Information Science and Technology*, Vol. 30 No. 4, pp. 205-214.

Bogers, T. and Koolen, M. (2017), "Defining and supporting narrative-driven recommendation", *Proceedings of the 11th ACM Conference on Recommender Systems, Como, ACM, New York, NY*, pp. 238-242.

Buchanan, G. and McKay, D. (2011), "In the bookshop: examining popular search strategies", *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries 2011, ACM, New York, NY*, pp. 269-278.

Eberhard, L., Walk, S., Posch, L. and Helic, D. (2019), "Evaluating narrative-driven movie recommendations on Reddit", *Proceedings of the 24th International Conference on Intelligent User Interfaces, ACM, New York, NY*, pp. 1-11.

Elsweiler, D., Wilson, M. and Kirkegaard Lunn, B. (2011), "Understanding casual-leisure information behaviour", in Spink, A. and Heinström, J. (Eds), *New Directions in Information Behaviour*, Emerald Group Publishing, Bingley, pp. 211-241.

Gäde, M. and Petras, V. (2016), "Hood or hypertext: a comparison of offline and online book search sessions", *CLEF*, pp. 1097-1105.

Goodall, D. (1989), *Browsing in the Public Libraries*, LISU Occasional Paper No 1, Library and Information Statistics Unit, Loughborough.

Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2010), *Multivariate Data Analysis*, Prentice-Hall, Upper Saddle River, NJ.

Hypén, K. and Mäkelä, E. (2011), "An ideal model for an information system for fiction and its application: Kirjasampo and semantic web", *Library Review*, Vol. 60 No. 4, pp. 279-292.

Koolen, M., Bogers, T., van den Bosch, A. and Kamps, J. (2015), "Looking for books in social media: an analysis of complex search requests", *Proceedings of ECIR'15, Springer, Cham*, pp. 184-196.

Kraaykamp, G. and Dijkstra, K. (1999), "Preferences in leisure time book reading: a study on the social differentiation in book reading for the Netherlands", *Poetics*, Vol. 26 No. 4, pp. 203-234.

Liu, C., Cole, M., Belkin, N. and Zhang, X. (2012), "Exploring and predicting search tasks difficulty", *Proceedings of CIKM'12, ACM, New York, NY*, pp. 1313-1322.

Lorigo, L., Pan, B., Hembrooke, H., Joachims, T., Granka, L. and Gay, G. (2006), "The influence of task and gender on search and evaluation behavior using Google", *Information Processing & Management*, Vol. 42 No. 4, pp. 1123-1131.

Miesen, H.W.J.M. (2003), "Predicting and explaining literary reading: an application of the theory of planned behavior", *Poetics*, Vol. 31 Nos 3-4, pp. 189-212.

Mikkonen, A. and Vakkari, P. (2012), "Readers' search strategies for accessing books in public libraries", *Proceedings of the 4th Information Interaction in Context Symposium, ACM, New York, NY*, pp. 214-223.

Mikkonen, A. and Vakkari, P. (2016a), "Finding fiction: search moves and success in two online catalogs", *Library and Information Science Research*, Vol. 38 No. 1, pp. 60-68.

Mikkonen, A. and Vakkari, P. (2016b), "Readers' interest criteria in fiction book search in library catalogs", *Journal of Documentation*, Vol. 72 No. 4, pp. 696-715.

Mikkonen, A. and Vakkari, P. (2017), "Reader characteristics, behavior and success in fiction book search", *Journal of the Association for Information Science and Technology*, Vol. 68 No. 9, pp. 2154-2165.

O'Connor, D. (1987), "Elders and higher education: Instrumental or expressive goals?", *Educational Gerontology*, Vol. 13 No. 6, pp. 511-519.

Oksanen, S. and Vakkari, P. (2012), "Emphasis on examining results in fiction searches contributes to finding good novels", *Proceedings of JCDL 2012, ACM, New York, NY*, pp. 199-202.

Pejtersen, A.M. (1989), "The bookhouse: modeling user's needs and search strategies as a basis for system design", Risø Report No. M-2794, Risø National Laboratory, Roskilde.

Pöntinen, J. and Vakkari, P. (2013), "Selecting fiction in library catalogs: a gaze tracking study", *Proceedings of the 17th International Conference on Theory and Practice of Digital Libraries, Vol. 8092, Springer*, pp. 72-83.

Ross, C., McKehnie, L. and Rothbauer, P. (2006), *Reading Matters*, Libraries Unlimited, Westport, CT.

Ross, C.S. (2001), "Making choices: what readers say about choosing books to read for pleasure", *The Acquisition Librarian*, Vol. 13 No. 25, pp. 5-21.

Roy, M. and Chi, M.T.H. (2003), "Gender differences in patterns of searching the web", *Journal of Educational Computing Research*, Vol. 29 No. 3, pp. 335-348.

Saarinen, K. and Vakkari, P. (2013), "A sign of a good book: readers' means of accessing fiction in the public library", *Journal of Documentation*, Vol. 69 No. 5, pp. 736-754.

Saarti, J. and Hypén, K. (2010), "From thesaurus to ontology: the development of the Kaunokki Finnish fiction thesaurus", *The Indexer*, Vol. 28 No. 2, pp. 50-58.

Schamber, L. (1994), "Relevance and information behavior", in Williams, M. (Ed.), *Annual Review of Information Science and Technology*, Vol. 29, Information Today, Medford, NJ, pp. 3-48.

Spiller, D. (1980), "The provision of fiction for public libraries", *Journal of Librarianship and Information Science*, Vol. 12 No. 4, pp. 238-266.

Stockmans, M. (2003), "How heterogeneity in cultural tastes is captured by psychological factors: a study of reading fiction", *Poetics*, Vol. 31 No. 4, pp. 423-439.

Tang, M.-C., Sie, Y.-J. and Ting, P.-H. (2014), "Evaluating books finding tools on social media: a case study of aNobii", *Information Processing and Management*, Vol. 50 No. 1, pp. 54-68.

Thudt, A., Hinrichs, U. and Carpendale, S. (2012), "The Bohemian bookshelf: supporting serendipitous book discoveries through information visualization", *Proceedings of CHI'12, ACM, New York, NY*, pp. 1461-1470.

Usherwood, B. and Toyne, J. (2002), "The value and impact of reading imaginative literature", *Journal of Librarianship and Information Science*, Vol. 34 No. 1, pp. 33-41.

Vakkari, P. (2002), "Subject knowledge, source of terms and terms selection in query expansion: an analytical study", *Proceedings of the 24th ECIR, Springer, Berlin*, pp. 110-123.

Vakkari, P. and Pöntinen, J. (2015), "Result list actions in fiction search: an eye-tracking study", *Proc. of JCDL'15, ACM, New York, NY*, pp. 7-16.

Wildemuth, B. (2004), "The effects of domain knowledge on search tactic formulation", *Journal of the Association for Information Science and Technology*, Vol. 55 No. 3, pp. 246-258.

Further reading

Mikkonen, A. and Vakkari, P. (2015), "Books' interest grading and fiction readers' search actions during query reformulation intervals", *Proceedings of Joint Conference on Digital Libraries 2015, ACM, New York, NY*, pp. 27-36.

Corresponding author

Pertti Vakkari can be contacted at: pertti.vakkari@tuni.fi