Information Processing and Management 53 (2017) 905–922

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Evaluating user search trails in exploratory search tasks

Chathra Hendahewa[a,c,*], Chirag Shah[b]

[a] Department of Computer Science, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA
[b] School of Communication and Information, Rutgers University, 4 Huntington Street, New Brunswick, NJ 08901-1071, USA
[c] Detectica LLC, 84 Wooster Street, Suite 703, New York, NY 10012, USA

ARTICLE INFO

Article history:
Received 10 October 2016
Revised 26 March 2017
Accepted 1 April 2017
Available online 12 April 2017

Keywords:
Exploratory search
Evaluation
Search trail

ABSTRACT

Exploratory search is characterized by a user's uncertainty towards a complex information seeking task. A user conducting such a search in an information retrieval (IR) system may need help and recommendations that are beyond mere query suggestions. In this paper we propose a new method for recommending search trails to struggling users. We first use a search process prediction model from the literature to predict whether a user is likely to under-perform in an exploratory search task, and given that case, recommend a search trail based on other users' search behaviors in a similar context. We then present a method to evaluate the effectiveness of these recommendations that involves two different evaluation criteria. First, we use Open Directory Project (ODP)-based categorization of user-traversed Web pages to evaluate each user's information coverage. Next, we evaluate the order of users' search trails while simultaneously incorporating a novel set of metrics that use adjacency of queries issued and Web pages traversed. To evaluate search trails, we incorporated proposed metrics with transactional log data from multiple user studies in which more than 300 users conducted exploratory search tasks on different topics.

We demonstrate the effectiveness of the proposed evaluation criteria by measuring how the recommended search trails lead to improvements in both information space coverage and search performance metrics for users across multiple user search datasets. Based on the analysis results, we demonstrate that the order of the recommended search trails plays a significant role and it outperforms the random order of search trails thus being beneficial for the struggling users in improving their overall search effectiveness. We also show that by providing search trail recommendations, users are able to discover more information across multiple facets (in breadth) as well as investigate certain facets in more detail (in depth). These findings provide substantial evidence across multiple datasets to confirm that recommended search trails improve users' information seeking coverage and overall knowledge acquisition throughout their search processes.

1. Introduction

Exploratory searching behavior is characterized by significant uncertainty towards the goals of a search or a dearth of knowledge about a search topic's domain (White and Roth, 2009, pp.10; Kuhlthau, 2004). As a consequence, searchers who are exploring need support to help them sift through the unknown. One possible solution is to help exploratory searchers

* Corresponding author.
E-mail address: chathra.hendahewa@gmail.com (C. Hendahewa).

http://dx.doi.org/10.1016/j.ipm.2017.04.001
0306-4573/

by automatically providing them with *trails*, or previously traveled search paths that have proven useful to other searchers with similar exploratory search behaviors. Pirolli and Card (1999) developed a sophisticated model of user behavior called *information foraging* derived from animals' behavior when foraging for food in the wild. They use a foraging metaphor to discuss how information seekers could use cues left by previous visitors to find and consume patches of information in a collection to satisfy their own needs. We utilize both the information foraging theory and a recent prediction and recommendation approach that incorporates implicit features proposed by Hendahewa and Shah (2015) to recommend search trails to users to improve their overall information coverage and search performance.

When analyzing user search behavior, developing a user profile is a challenging task because it may change throughout the search process. For example, over time a user can develop from an ineffective searcher to an effective searcher or vice-versa. Evaluating a user's profile in a dynamic manner in order to provide real time recommendations to improve their search performance can be considered an improvement upon traditional recommender systems. Current IR systems often use query auto-completion and related services to provide recommendations. Considering that some exploratory searchers may not even have a reasonable query to begin with, these methods can be limiting. Rather than adhere to a traditional recommender system's paradigm, the work presented in this paper will attempt to evaluate recommendations based on dynamic user behavior analysis throughout the search process in a specific order of search paths and visitations.

1.1. Research objective

In this article, we focus on addressing two main research objectives related to recommendation and evaluation of search trails.

- **RQ1**: Can we assist struggling users by recommending them better search paths in order to improve their search performance and reach the task goal?
- **RQ2**: Provide a set of metrics for performance assessment based on the recommended search trails.

It should be noted that for the purpose of this article, we use the definition of *search trail* as being defined by White and Huang (2010) as a search path that begins with a search engine query and comprises a set of pages visited until the trail terminates with a new query or an inactivity or timeout.

For users who are deemed to be under-performing according to the evaluations proposed in the algorithm by Hendahewa and Shah (2015) as shown in the pseudo-code below, we provide similar interventions in order to improve their search process (Algorithm 1).

```
Data: Search trails of each user
Result: Simulated search path recommendations for under-performing users;
while Remaining under-performing performers(l) do
    Find matching high performers with similar first search path;
    for (h = 1; h ≤ num(high performers); h++) do
        if QueryEditDist(Firstsearchpath_l - Firstsearchpath_h) ≤ Threshold
        then
            Switch remaining searchpath_l with remaining searchpath_h;
        else
            No simulation;
        end
    end
end
```

Algorithm 1: Search Process Recommendation Simulation Algorithm.

Many possible methods of search intervention have been considered, including query suggestions (Karisani, Rahgozar, & Oroumchian, 2016; Marchionini & Shneiderman, 1988; Niu & Kelly, 2014; Vidinli & Ozcan, 2016), query auto-completion (Cai & de Rijke, 2016; Smith, Gwizdka, & Feild, 2016), trail suggestions (Singla, White, & Huang, 2010; Wexelblat & Maes, 1999; White, Bilenko, & Cucerzan, 2007) and entity suggestions (Balog, Bron, & De Rijke, 2011; Brando, Santos, Ziviani, de Moura, & da Silva, 2014). In exploratory search, where the emphasis is placed on complex extended search processes, we argue that the recommendation of search trails is useful and appropriate. Recommendation of search trials goes beyond both related query suggestions and relevant document suggestions, as search trails allow a user to explore diverse and undiscovered areas of relevant information. In assessing **RQ1**, by providing search trail based recommendations (that were deemed to be successful by other searchers) to the struggling users we evaluate whether they would benefit from such recommendations.

Further, we propose a set of metrics that utilizes information coverage using Open Directory Project (ODP)-based categorization of Web pages and also proposed measures to evaluate goodness of a search trail that address **RQ2**. Using search trails as the unit of analysis and search trail recommendation as the mode of intervention in real time, the work reported in this paper will demonstrate that we can improve search performance metrics and achieve higher information space coverage.

1.2. Contributions

The major contributions of this article in relevance to the research objectives stated above are as follows.

- Evaluate the effect of suggested search trails (based on recommendations following Hendahewa & Shah (2015)) on the coverage of information (sub topics/categories) during the search task.
- Evaluate the goodness of the suggested search trails based on the execution order.

The rest of the paper is organized as follows. In the next section, we discuss the relevant literature on information seeking models, exploratory search, and evaluations of search trails. This is followed by the proposed evaluation metrics to assess search trails using the information space coverage and the order of search trails. The experiments section presents the data used in the analysis and the results obtained in detail. The findings from the proposed evaluation of search trails, the relevance to the IR community, and the limitations are presented in the discussion section. Finally, we conclude the paper with a summary and suggestions for future work based on the proposed line of research.

2. Background

2.1. Information seeking models

Understanding the theories behind information seeking models would enable one to carefully evaluate what types of search behaviors could lead to more effective information seeking strategies that yield relevant information.

Borgman, Hirsh, and Hiller (1996) stated that “The ability to study online searching behavior in all its complexity requires research methods that can capture the details of the search process, rather than just the product or output of a search session”. They suggested that building quantitative models of user search processes based on transactional log data would better capture user search behavior. The Information Search Process (ISP) model by Kuhlthau (2004, 1991) describes the process of information seeking to accomplish a task’s goal from a user’s perspective with regard to six stages: initiation, selection, exploration, formulation, collection, and presentation. The types of search behavior that are exhibited within each of these stages differently describe how search processes evolve during an information seeking task. Belkin’s Anomalous State of Knowledge (ASK) model (Belkin, Oddy, & Brooks, 1982) explains how a person’s information need, which arises due to some sort of uncertainly in their knowledge base, can be fulfilled through information seeking behaviors that change their initial state of knowledge. The berry-picking model of information seeking introduced by Bates (1989) explains how a user’s query may shift as they absorb the information gathered throughout their search process. She discusses how information seen at one point in the search process might lead the search into a new, unanticipated direction that still relates to the initial information need.

Pirolli and Card’s (1999) information foraging model explains how information paths and cues left by previous searchers could help new seekers more effectively traverse the information seeking process. Later, Pirolli and Fu (2003)developed and validated computational cognitive models of Web navigation behavior based on information foraging theory. Each time a search path is traversed, an “information scent” leaves important clues as to how a user can utilize their search strategies to find useful and relevant information for their search tasks.

Downey, Dumais, and Horvitz (2007) proposed a framework for language of search activity models in which activities such as session initiation, query execution, SERP (Search Engine Result Page) click, non-SERP click, and session end were incorporated using a state space model.

2.2. Exploratory search and search recommendation

Exploratory search is a form of information seeking that describes the activity of attempting to obtain information through a combination of querying and collection browsing (White and Roth, 2009, pp.10). Unlike fact-finding or known item search, exploratory search usually spans across many facets and sub tasks that require multiple queries and search sessions before an information searcher can find meaningful and useful information that extends their knowledge about the task at hand.

As suggested by several researchers such as Bawden (1986), Foster and Ford (2003), Budd (2004), White and Roth (2009, pp.10) and Marchionini (2006), exploratory search includes abundance of behaviors beyond *regular* search behavior, such as exploration, uncertainty, creativity, knowledge discovery, and learning. These unique aspects make exploratory search into a complex search process that involves multiple search trails across multiple facets that a user has to investigate in order to fulfill their goals. There has been recent attempts to further analyze how exploratory search is different from look-up search tasks such as fact-finding, question answering has shown that systems can distinguish the task types using different characteristics such as task time, scroll depth, etc. of a search session (Athukorala, Glowacka, Jacucci, Oulasvirta, & Vreeken, 2016).

The traditional recommender systems (Ricci, Rokach, Shapira, & Kantor, 2010) have focused on providing recommendations to users based on their profiles. In the past, most researchers (e.g. Basu, Hirsh, & Cohen, 1998; Cai, Wang, & de Rijke, 2016; Miranda et al., 1999; Park, Kim, Choi, & Kim, 2012) have studied new approaches that go beyond such traditional recommendation systems to effectively provide personalized recommendations to users with the help of data mining techniques.

Recently, many researchers have investigated how mining and evaluating log data on users' search satisfaction can provide personalized recommendations (He, Qvarfordt, Halvey, & Golovchinsky, 2016). Many attempts have been made to model user search behavior on a global level applicable to all users. The major drawbacks of such user behavior models, as Hassan and White (2013) point out, are that they fail to distinguish between individual preferences and personal satisfaction in searching, and only cater to a larger audience. This emphasizes that a personalized approach to search behavior analysis is important in order to develop tailored models that provide recommendations to satisfy each individual user. Hassan and White (2013) show that performing personalization at both the individual user level and the cohort level (where individual users were grouped into cohorts based on topical similarities between user interests) using query, session, Search Engine Results Page (SERP), and pattern of behavior features, plus the global level model, leads to better document retrieval accuracies (based on precision and recall) and better F-scores of results.

Understanding the user is a major component in any system that involves interaction with a human user. Due to the cognitive abilities and the highly individualistic traits of each user, understanding individuals when interacting with a system enables the system to adapt to its user's need and provide a better interaction experience (White & Roth, 2009). In online search systems, this is a major concern because, by understanding the user, search recommendations can be better personalized, better focused and also catered to specific needs. In order to facilitate search systems' ability to understand users, many researchers have taken two viewpoints - one to provide sequential pattern analysis and the other to build behavioral models. Most of these cases have been applied to general information search, but not specifically to exploratory search where users are dealing with high levels of uncertainty. This makes the problem more complex and challenging, but also quite interesting because exploratory searchers' behavior patterns can enable them to mitigate issues such as uncertainty, lack of focus, and confusion, thereby helping them achieve their goals.

2.3. Search trail recommendation

White and Huang (2010) note that search trails can be effectively evaluated to measure diversity, coverage, and novelty, all of which are important for exploratory search. Further, Hassan Awadallah, White, Pantel, Dumais, and Wang (2014) analyzed how the search trail based recommendations could improve search performance of users.

Olston and Chi (2003) proposed an user interface called *ScentTrails* that utilizes the context of information foraging theory and highlights hyperlinks to indicate paths to search results. This interface and related user studies also showed that by allowing the users the take advantage of both searching and browsing to locate content matching complex information goals using the concept of *trails* or *search paths*, the users would benefit in achieving the task goals more effectively.

Singla et al. (2010) demonstrated the value in search trails by highlighting interesting differences in the performance of various trail-finding algorithms and how one can find best search trail that outperforms others for a given query.

Trail selection methods could discount trails with numerous cases of rapid backtracking, or they could maximize relevance, coverage, diversity, novelty, and utility by recommending only high quality trails. Alternatively, we can personalize trail recommendation by weighting trails based on users' re-finding behavior. We can also perform a-priori trail analysis to recommend trails when the destination is unclear (i.e., users end up on many pages), and present trail destinations when the destination is clear (i.e., many users end up at the same page).

3. Performance assessment in exploratory search

Given the complex nature of the exploratory searches tasks conducted on the Web, assessing the performance is a challenging task. Egusa et al. (2010) proposed a user-centered method to evaluate the performance of an exploratory search task by comparing the users' mental representation of the topic using concept maps. They were able to show empirically that the concept maps between the pre- and post-searches indicated that the users significantly changed their knowledge structure of a topic by completing the exploratory search task.

Since exploratory search tasks are multi-faceted (with multiple search sub tasks to cover the entire task goal), users have to explore a topic's multiple sub areas to successfully satisfy their information needs. They can do so via observable behaviors, such as conducting multiple search queries, query reformulations, multiple page visits, and longer search sessions. Recent attempts in evaluating exploratory search systems using sense-making framework shows that users utilized the search to find structure and ideas instead of just accumulating information facts when conducting exploratory search (Qu & Furnas, 2008).

Incorporating the major points of concept maps and sense making frameworks that allow us to evaluate what users have gained by conducting the exploratory searches, we use a more search feature oriented approach. In order to measure and evaluate how users are performing in an exploratory search task and to identify whether users need intervention to improve their search process, we use a predictive model proposed by Hendahewa and Shah (2015) based on an extracted set of implicit features that indicate the major aspects of exploratory search. Further, the evaluation metrics that we propose combines both the user queries and information retrieved (as Web pages) in evaluating how well the users conduct the search task and also uses the idea of concept maps in a higher context using topic and sub topic coverage.The two methods we use to evaluate the *goodness* of recommended search trails, as listed below.

- Information space coverage based on the organization of the sub topics using DMOZ open Web portal of Web site directories as used in Hassan Awadallah et al. (2014) (http://www.dmoz.org) to evaluate the performance of user search trails.
- Order of the execution of different search trail measures through the sequence of queries and Web pages.

3.1. Evaluating information space coverage

To evaluate the overall performance of pre- and post- search trail recommendation information seeking, we use information space covered over the areas of sub topics/categories. Using simulations, we demonstrate that our search trail-based recommendations improve users' search performance based on information coverage according to the Open Directory Project (ODP) categorization of Web pages across depth and breadth of the resulting tree structure.

Once the recommendations were performed based on the *Search Process Recommendation Simulation Algorithm* explained in Hendahewa and Shah (2015), we analyze the effect of those recommended trails at different points over the search process, such as the end of the first, second, and third search trail. This analysis shows how the recommended search trails could affect the information coverage for a particular topic. Once the recommended search trail is generated, each of the corresponding Web pages in the actual and suggested search paths are categorized using the DMOZ ODP-based categorization (http://www.dmoz.org). It should be noted that the Web pages that did not have ODP categorizations in their repository were manually labeled adhering to the category/sub category naming guidelines based on their content. Next, the depth (maximum number of edges from top node to leaf node) and the breadth (maximum number of edges originating from top node) of the categorization tree are calculated for each recommendation.

3.2. Evaluating order of search trails

To assess the order of recommended search paths and their order of recommendation, we propose three different metrics that consider query order and page order, as described below.

As the first measure, we incorporate the Likelihood of Discovery (LD) measure for each Web page as shown in Eq. (1).

$$LD(p) = \frac{|C(p)|}{|C|} \tag{1}$$

where $C = \{c | c \in Web\_page\_clicks\}$ and $C(p) = \{Clicks\_on\_page\_p\}$.

Therefore, the LD measure for each user u over time t can be calculated as shown in Eq. (2).

$$LD(u,t) = \sum_{p=1}^{|C(u,t)|} LD(p) \tag{2}$$

where $C(u,t) = \{c | c \in C, clicked\_by\_u, 0 \le t \le n\}$.

Thus, this measure calculates how easy/difficult it was for a user to find a page. A high value indicates easy findability, while a low value indicates difficult findability and low user traffic. For example, most users find easy pages (e.g., Wikipedia pages, or general information pages) early during the search process and difficult pages later during the search process.

As the second evaluation measure, we use the concept of adjacency in evaluating recommended search trails by following the approach mentioned in He et al. (2009). Given a query q, this method uses the rank based on an ordered list of queries (q_1 to q_m) that immediately follows q from the total corpus of queries in the data set based on co-occurrence of the total corpus.

For each recommended search trail, we evaluate the rank of the next query. We also weight each trail's total number of clicks, which indicate its goodness based on query order. This is similar to the *Borda Count* measure (Saari, 2000) that assigns a higher weight to items that are ranked higher in a list as shown in Eq. (3).

$$Goodness(q, q_i) = \frac{1}{Adj(q, q_i)} * |C(q_i)| \tag{3}$$

where $C(q_i) = \{c | c \in C, clicks\_from\_q_i\}$ and $Adj(q, q_i) = i$.

The last measure tries to evaluate the order of Web page visitations. Given a Web page p, this method uses a ranked list of Web pages (p_1 to p_n) that is followed by p from the total corpus of Web pages ranked according to their co-occurrence.

For each recommended search trail, we evaluate its goodness using the new ranked metrics of Web pages, which corresponds to the total set of Web pages in that search trail multiplied by the number of times users found them useful. Here the *usefulness* is measured based on the widely used notion of dwell time greater than or equal to 30 s (Fox, Karnawat, Mydland, Dumais, & White, 2005) as shown in Eq. (4).

$$Goodness(p, p_i) = \frac{1}{Adj(p, p_i)} * |UC(p_i)| \tag{4}$$

where $UC(p_i) = \{c | c \in C, time(p_i) \ge 30s\}$ and $Adj(p, p_i) = i$.

4. Experiments and user studies

The experiments were conducted on transactional log data collected from three different user studies; (1) Lab study (2) Classroom based study (3) Live study.

All the user studies used open calls to recruit undergraduates from major universities in different countries across various majors and classes. All the participants were compensated with either cash, gift cards, or class credit to motivate their willingness to participate. All three studies involved users working on an exploratory search task to find information on a specific topic using online search and compiling a report on that specific topic. The participants were not given advanced notice of their topics to ensure limited bias towards prior knowledge acquisition.

The details of each user study is explained as follows.

4.1. Lab study

This section describes two user studies conducted in a lab setting where a where a researcher monitored/guided sessions while the participants performed the exploratory search tasks.

4.1.1. Lab study 1: Global warming

In the first study, the participants were students recruited from a major US university through open calls that were distributed through various email-lists. Participants signed up for the study through an online form. Of the 68 participants (12 individuals, 10 dyads, and 12 triads) that were recruited, 40 were female and 28 were male, with ages ranging between 18 and 24. Most of the participants (60%) reported using Windows operating system. Moreover, all of the participants indicated having intermediate to advanced search skills. It should be noted that in the analysis for this article, we only considered the 12 individual users' search processes since we are focusing on individual user search performance and excluding the rest of the participants as they conducted the task in a group setting.

The participants were asked to collect relevant information in an exploratory search task designed to simulate a realistic work-task. Logs for each user correspond to approximately 30 min of active search. The participants were compensated using gift cards for their participation in the study. The topics provided for the exploratory search task are as follows.

Task Description: A leading newspaper agency has hired your team to create a comprehensive report on the causes, effects, and consequences of the climate change taking place due to global warming. As a part of your contract, you are required to collect all the relevant information from any available online sources that you can find. To prepare this report, search and visit any website that you want and look for the specific aspects listed in the guideline below.

As you find useful information, highlight and save relevant snippets. Later, you can use these snippets to compile your report, which will be no longer than 200 lines.

Your report on this topic should address the following: Description about global warming, how it affects climate change, scientific evidence about global warming affecting climate change, causes of global warming, consequences of global warming causing climate change, measures that different countries around the globe have taken over the years to address this issue, and recent advancements in addressing this issue. Also describe different viewpoints people have about global warming (specify at least three different viewpoints) and relate those to the topics' controversies.

4.1.2. Lab study 2: Quiet revolution

The second lab study was also identical to first lab study explained above in all the settings such as in recruiting undergraduate students and the participants were asked to collect relevant information on a given exploratory search task with an active search time of 30 min. The only difference was in the topic assigned (Quiet Revolution) and the users had to perform a more comparative task by comparing different revolutions that took place in the past. The number of individual participants in this user study was 37 users and out of which we had to eliminate data of two user sessions due to a technical error in capturing the data correctly. The topic given for this user study is shown below.

Task Description: As a history buff, you have heard of the Quiet Revolution, the Peaceful Revolution, and the Velvet Revolution. For a skill-testing question to win an iPod you have been asked how they differ from the April 19th Revolution.

Search and visit any Website that help you find information on this topic. As you find useful information, highlight and save relevant snippets. Make sure you also rank your snippets based on their quality and usefulness (Shah, Liu, González-Ibáñez, & Belkin, 2012).

4.2. Classroom study

This user study data comprised a large number of users (308) who worked on a prompt on the same topic of red yeast for 45 min. This was conducted within a classroom setting where all the participants were undergraduate students who were part of a specific class in a major foreign university in Netherlands. The entire study was conducted as part of a classroom assignment in English and the participants used English as their main language when conducting this study. The lab moderators took many precautions as possible to ensure that the participants use English search engines instead of non English ones, by redirecting them to Google UK, if they tried to use other search engines (via the search system) and provided precise instructions as to only user English search engines for the course of this study. The users were able to receive more guided help by lab moderators while performing their task. These users were also not informed of their topics in advance. Additionally, the system was explained to users, and they were given a 5 min warm-up task to get used to the system and practice. The participants were compensated for class credit for participation in the study.

The task description provided to the participants is as follows.

- Red Yeast Rice: For this task, you will be researching the safety of Red Yeast Rice.
  The official has heard that French officials have raised some concerns about the safety of Red Yeast Rice and potential contamination, and would like a briefing on its potential risk.
  Your task is to act as an advisor to an official within the science ministry. You are advising an official on the issues below. The official is not an expert in the area, but you can assume they are a generally informed reader. They are interested in the best supported claims in the documents. Produce a summary of the best supported claims you find and explain why you think they are. Note you are not being asked to create your own argument or summarize everything you find but rather, make a judgment about which claims have the strongest support.

4.3. Live study

This study was a live study where a total of 18 participants performed the task at their convenience on their own personal computers by accessing a web-based search tool (Hendahewa & Shah, 2015).

Target participants were undergraduate students who are highly fluent in the English language with intermediate to advanced online search and typing skills. We targeted undergraduates from majors such as Computer Science, Information Technology, Informatics and Journalism studies through open calls that were spread through various email-lists. Through an online registration form, the participants signed up for the study.

The user study system was designed and developed as a web application and a browser add-on that can be accessed online. The system facilitates the automatic progression of log in, questionnaires, video tutorials on how to use the system, practice session, pre-task quiz, topic selection, searching information online to complete the exploratory search task, post-task quiz, and questionnaires. Since the browser plug-in is customized for the Firefox browser, so we instructed the participants to install Firefox before accessing the user study system. Participants had the opportunity to select two out of five topics to perform two exploratory search tasks over two separate sessions. There were 12 participants who selected technology task and only 9 participants selected the health task as one of their two options. The other 3 topics (Entertainment, Arts/History and Environment/Energy) was selected by fewer participants leading to sparse data so was omitted from the analysis for the purpose of this paper.

Participants were compensated with Amazon gift cards for participation in the study. The exploratory task was allocated a minimum of 1 h and the participants were given a maximum of 2 h to complete each task.

Descriptions of the exploratory task topics are explained below.

- Technology: You are a journalist for a national newspaper. With the recent focus on security breaches and consumer data breaches, vulnerabilities in data and software have become an important topic of interest to a multitude of people and organizations ranging from government, companies, and technical industries to the general public. You have been asked to write an article about data and software vulnerabilities. Your article should include a basic introduction to the topic, along with past and present events. Your article should appeal to as many people as possible, including an unfamiliar, lay audience and people in affected businesses. It should cover different aspects of software vulnerabilities that are relevant to their daily lives. The article should also focus on the measures taken to minimize these risks at industry, government and consumer levels.
  When conducting this task, you should collect as much information as you can by searching online. You can use snippets to collect information that you deem useful for writing the report and copy them into your article. Nevertheless, the article should be well-written with a proper flow so that its readers would be able to understand the context and gather much information about this topic. Your article should be around 1000 words.
- Health: One of your close family friends talks to you about his current health and wellness issues. Because he is not tech-savvy, you are trying to help him find useful information. He needs to understand your report to get more information about his health and wellness requirements. Your family friend is a 30-year-old, type 2 diabetic male who wants to lose weight. However, he is usually very busy with work and family and can only spare 3 h each week to exercise. He has Internet access and an iPhone but cannot use them effectively. He has asked for your help. Assemble a diet and exercise program for him, including its benefits and risks. This report should also cover aspects of possible applications he could use to easily to monitor and control his wellness. When conducting this task, you should collect as much information as you can by searching online. You can use snippets to collect information that you deem useful for writing the report and copy-paste them into your article. Nevertheless, the article's context should cater to your family-friend's needs, and it should elaborate on his various health-related options. Your report should be around 1000 words.

Table 1
Datasets used in experiments.

User study	Dataset (Topic)	# of Users (Actual)	# of Users (Used in experiments)
Live study	Technology	9	9
Live study	Health	12	12
Lab study 1	Global warming	12	12
Lab study 2	Quiet Revolution	37	35
Classroom study	Red Yeast Rice	308	238
Total		*378*	*306*

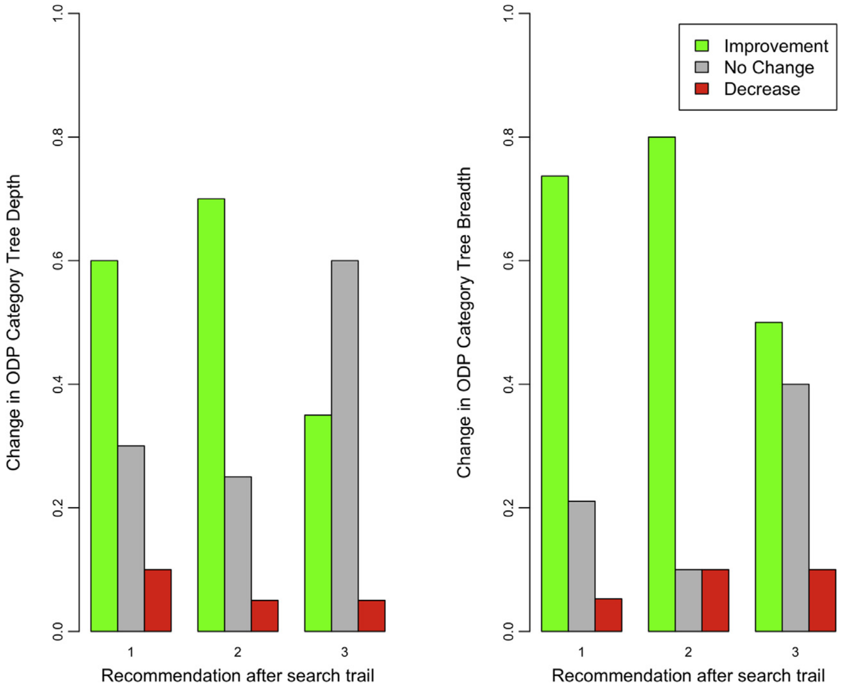

Fig. 1. Search trail recommendation effect for Quiet Revolution task.

5. Data and analysis

Based on the user study experiments explained in the Section 4, the datasets correspond to each exploratory search task topic. The collected transactional log data was comprised of timestamps, user search actions, queries, web pages, bookmarks, and snippets users considered relevant.

The datasets contain search logs of users performing exploratory search tasks on five different topics: (1) Technology (Hendahewa & Shah, 2015) (2) Health (Hendahewa & Shah, 2015) (3) Quiet Revolution (Shah et al., 2012), (4) Global warming and (5) Red Yeast Rice. It should be noted that because an analysis of a user search process-based recommendation is only relevant if we can observe search processes with at least 2 search trails, some of the users were discarded from the analysis due to data sparsity or data recording issues. The details of the datasets along with the respective user studies that are used in the analysis are shown in Table 1.

The data analysis section aims to focus on evaluating the following goals in line with the research questions specified in the Introduction.

- Evaluate if the recommended search trails assist the users to improve their search paths and achieve higher performance by comparing the ODP categorization based space coverage improvements. (In relevance to RQ 1).
- Evaluate the effectiveness of the set of metrics (Likelihood of discovery, goodness of search trail based on query order and Web page order) when compared to actual search trail orders. (In relevance to RQ 2).

5.1. Analysis of information space coverage

Fig. 1 shows the ratio of the effect of recommendations at the end of first, second and third search trails for the datasets - Quiet Revolution task. Additional figures on other datasets are shown in Appendix A that confirms that the results are consistent across different data sets spanning multiple user studies. It can be observed across all datasets that there is a

Table 2
Recommendation effect on ODP categorization after second search trail.

User study	Dataset	Percentage of users with improvement	
		in breadth	in depth
Live study	Technology	60%	70%
Live study	Health	60%	60%
Lab study 1	Global warming	45%	80%
Lab study 2	Quiet Revolution	70%	80%
Classroom study	Red Yeast Rice	60%	60%

substantial improvement in breadth of the information coverage categories that illustrates the recommended search trails are beneficial in improving the information space covered. This verifies that introducing new sub topics can act as a way of improving the exploration and knowledge discovery of users. It can also be observed that the depth of information covered is either improving or not making an adverse impact in a majority of the cases. This shows that by providing the recommendation the impact of finding information in depth is not adversely affected in most cases.

Another finding from these experiments is that the highest impact on improvement occurs (in both breadth and depth) if the recommendations were made at the end of first or second search trail. This indicates that early identification of struggling and recommendation of search trails leads to higher information gains. When the recommendation occurred after the second search trail, a higher percentage of users were able to gain more coverage across both breadth and depth dimensions of the ODP tree categorization on Web pages traversed as shown in Table 2.

The results provide evidence that the majority of the users were able to cover more sub categories within the particularly topic. This could be indicative of improved knowledge discovery and exploration, leading to enhancing the overall exploratory search process.

5.2. Analysis of order of search trails

Incorporating evaluation metrics to analyze how the order of the Web pages was decided using the metrics (Eqs. (1)–4) are described in Section 3.2. It should be noted that we are evaluating the recommended search trails for under-performing users.

5.2.1. Analyzing recommended search trails using likelihood of discovery (LD)

Using random order, actual order, and recommended order of search trails, we perform the analysis based on Likelihood of Discovery denoted by $LD(u, t)$ (Eq. (2)). Then we evaluate whether the recommended order of search trails yields better results that enable users to quickly find difficult information after fewer queries compared to the random and actual orders. LD results show that users have a harder time locating pages quickly when they follow their own search paths without recommendations.

It can be observed in Fig. 2 that if the queries and pages are recommended in a random order instead of following a specific search trail, LD fluctuates unlike the steadily decreasing trend (finding more and more difficult pages as opposed to easy pages) observed for recommended order of search trails. Although, the random order has a low LD at the end of second search trail, it is not consistent as the recommended search trail order in reducing the LD in subsequent steps. Also, it can be observed if the actual search trail order was considered the LD is higher in all search trails than the recommended order.

When analyzing LD at different points in time (instead of actual search trails), we can observe that the recommended search trails provide lower Likelihood of discovery, indicating that recommendations lead to finding more difficult pages (identical to Hendahewa & Shah, 2015) when compared to original search trails. This suggests that with fewer queries, users are able to find hidden but, relevant pages earlier in the search process. These observations are illustrated in Fig. B.7 in Appendix B.

At the end of third search trail the calculated Likelihoods of Discovery (LD) are shown in Fig. 3 for the Technology dataset. Further figures showing how the LD varied consistently at the end of fourth and fifth search trails are shown in Appendix B under Figs. B.8 and B.9. These figures illustrate that recommended order of search trails surpasses both random and actual order of search trails in terms of LD measure. Again, this emphasizes that the recommended search trail order ensures hard-to-discover pages are found first.

5.2.2. Analyzing recommended search trails using order of queries

Using $Goodness(q, q_i)$, we evaluate how well each recommended search trail would perform compared to actual and random search trails, as shown in Table 3. Each column shows the mean and standard deviation of the metric for different evaluations. We evaluate $Goodness(q, q_i)$ based on users' most frequently issued adjacency of queries. Table 3 demonstrates that across all datasets, the recommended order of search trails based on $Goodness(q, q_i)$ yields the best results. This shows that the order in which the queries are executed has an impact on users' ability to find better relevant information for their search tasks. For example, executing queries that are generic at the beginning of a search spurs exploration and causes an entire information seeking process to expand as each query iteration stems from previously discovered information. This

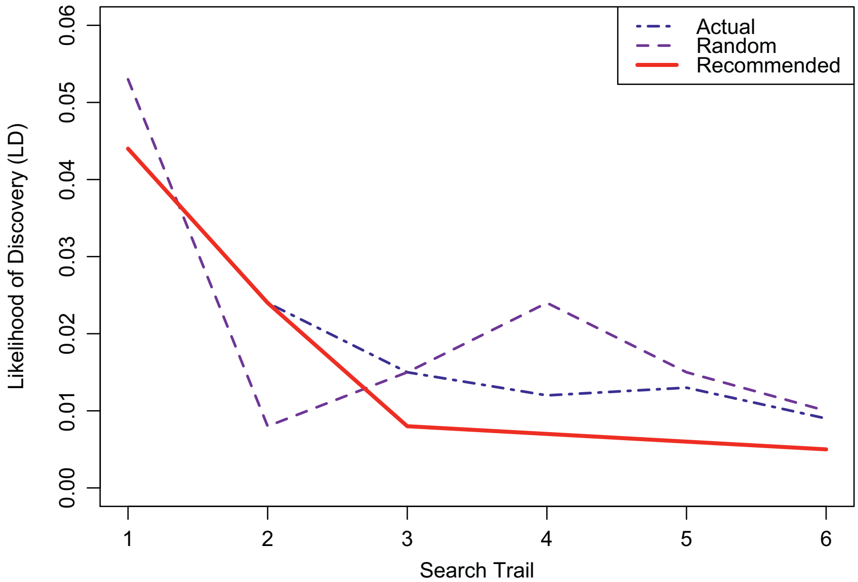

Fig. 2. LD with actual, random and recommended order at the end of each search trail.

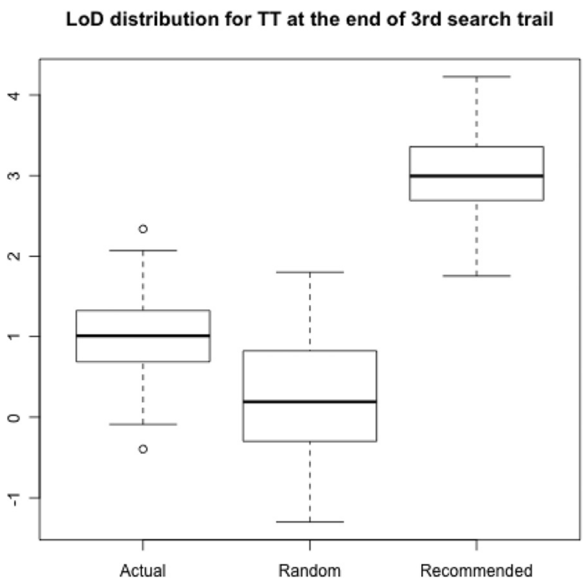

Fig. 3. *LD* comparison for 3rd search trail for technology data set.

means that once a user has gathered information throughout their search processes, sites that have the most clicks would yield a higher goodness measure when their information is deemed useful by an increasingly better informed user.

5.2.3. Analyzing recommended search trails using order of Web pages

We evaluate each search trail for recommendation using $Goodness(p, p_i)$ and evaluate how well it would perform compared to the actual search trail and a random order of search trails. Then we evaluate how the search trail effectiveness compares to a random order of Web pages for each of the datasets in consideration. Fig. 4 shows the distributions of the measure $Goodness(p, p_i)$ for the actual, random and recommended order of Web pages at the end of different points during the search process for the Technology dataset. Further results are shown in Appendix C for other datasets used in the analysis.

If the users were to follow their original order of visitations as shown in the actual search trails, they would do better than those that utilize random orders in most cases. However, they would definitely benefit from recommendations from high-performing users, as depicted by the recommended distribution. This observation of recommended order of search trails outperforming random and actual orders of search trails is visible across all the datasets as depicted in Figs. C.10–C.13.

Table 3
Search trail goodness evaluation.

User study	Dataset	*Goodness*(q, q_i)	4th trail	5th trail
Live study	Technology	Actual	1.5± 0.2	2.0± 1.0
		Recommended	2.5± 0.2	3.0± 0.3
		Random	2.0± 0.5	1.5± 0.5
Live study	Health	Actual	1.3± 0.2	1.5± 0.2
		Recommended	2.5± 0.5	2.0± 0.5
		Random	1.0± 0.5	1.0± 0.5
Lab study 1	Global warming	Actual	2.0± 0.3	1.5± 0.5
		Recommended	2.5± 0.3	2.0± 0.3
		Random	0.8± 0.5	1.0± 0.5
Lab study 2	Quiet Revolution	Actual	1.0± 0.3	1.0± 0.2
		Recommended	2.0± 0.5	2.0± 0.3
		Random	0.8± 0.5	1.0± 0.5
Classroom study	Red Yeast Rice	Actual	1.3± 0.3	1.2± 0.3
		Recommended	1.5± 0.5	2.0± 0.3
		Random	1.0± 0.5	1.0± 0.5

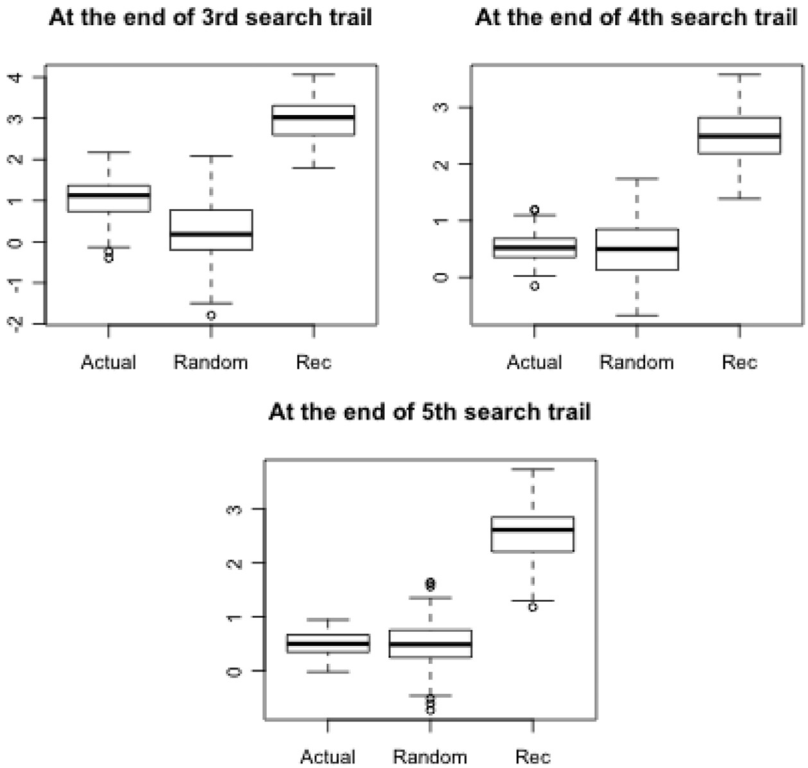

Fig. 4. *Goodness*(p, p_i) comparison for dataset: Technology.

5.3. Summary of results

The results are consistent across all the datasets collected during the different type of users studies explained in this article thus, shows the results can be generalized to show that recommended search trails outperform the actual search trail and benefit the users to gain better search performance (in meeting our first evaluation goal).

Based on the overall evaluation, we can summarize, that based on all three evaluation measures proposed (Likelihood of discover, Goodness of search trails based on query order, based on Web page order), the recommended search trails yield the best results when compared with the actual order and the random order that meets our second evaluation goal.

6. Discussion

We showed that the information coverage and the order of the recommended queries and pages at the end of each search trail are more important than merely providing a set of useful or relevant Web pages to the under-performing users.

According to the results and analysis presented in the previous section, we can observe that by providing under-performing users with better search trail recommendations that are developed by successful exploratory searchers, the former will discover more information across multiple facets (in breadth) as well as investigate certain facets in more detail (in depth). These findings provide substantial evidence across multiple datasets to confirm that recommended search trails improve users' information seeking coverage and overall knowledge acquisition throughout their search processes.

In this paper, search trail evaluation and the methods users employ to find information that meets search task goals are our major focuses. To this end, we clearly showed that providing under-performing users with the specific order of pages visited by proficient users has a more positive effect on search exploration compared to randomly ordered search trails and struggling users' actual search trails. This notion of order aligns with the information foraging theory (Pirolli & Card, 1999) where the user's current position and the scent or trail they have covered affect their proceedings. Further, in the analysis we found that the order in which search trails are executed within the search process is of utmost importance. Providing a mere set of query or page suggestions would not help unless they are ordered and shown to the user in their proper sequence. For example, recommending a search trail that corresponds to a query that uses "AT&T breach" as the first search iteration would not be useful until the user has discovered what types of data and software vulnerabilities were involved and which companies were affected (this applies to the Technology task). As proposed in this paper, search trails' "goodness" considers the order of search trail execution and determines which paths are more likely to provide users with a better search experience. Such search trail order evaluation metrics would enable exploratory search systems and search trail recommendation systems to evaluate how well different search trails would perform within various ordered sequences, thereby uncovering the best and most thorough paths that provide maximized information seeking efficiency.

Although substantial evidence corroborates our ability to improve the depth and breadth of discovered information via recommended search trails, this study is not without limitations. The main limitation of this work is that we assume a-priori that a user is conducting an exploratory search task in a specific topic and thus will have access to a set of other users' previously traversed search trails. We also assume that users will explicitly follow recommended search trails to improve their information seeking process. However, this assumption may not be realistic.

The proposed metrics and evaluation of recommended search trails would be highly applicable for most exploratory search tasks that focuses on a specific topic that spans multiple facets. Specifically, in the context of scientific literature search, or general information search about a certain topic that entails extensive browsing, key word search, sub topic identification, the proposed recommendations as means of search trails would be highly beneficial. With the proposed search trail recommendations and evaluations, the users would be able to find undiscovered search trails that lead to further information about a certain area/sub-topic they have not yet considered leading to both greater coverage of breadth (in terms of sub topic coverage) and depth (to go further a certain search path).

The proposed search trail recommendations can be incorporated in to a user exploratory search system in the context of providing recommendations as well as real time evaluation metrics to the user to get an understanding of how he/she is performing in meeting task goals. In addition to the search browsing interface, a secondary section could provide the recommended search trails as a trail map that highlights the source page, query suggestion and destination page along with the main key words that would be found by following that path. If the user opts to select such recommended trail, the evaluation metrics such as goodness of this search trail, would be presented to the user to provide live feedback.

One of the major obstacles when evaluating exploratory tasks spanning multiple user studies and topics is to evaluate the task difficulty. We tried to ensure that the tasks were all exploratory search tasks with topics spanning general topics such as health, environment, news, rather than requiring highly specialized levels of domain expertise that would enable users to successfully find information at a similar difficulty level. We conducted some pilot studies using undergraduates to see the difficulty level of the topics specifically in the lab studies by evaluating if the instructions given and the time to complete the task were sufficient across different topics and found no significant differences. We also asked pre and post questionnaires about the task difficulty from users for each session on a 5 point scale and most users confirmed the tasks to be of average difficulty level.

One of the complex issues that would arise in a real time recommendation and evaluation system is the evaluation of the user expertise in the topic and the user learning rate to be able to provide more refined and appropriate recommendations. In this paper, we have not considered the expertise level of a user but, assume based on the user studies we have conducted all involving undergraduate students and the similar level difficulty in task generation that the expertise level of users are not substantially different. Nevertheless, in a real time system, evaluating the user expertise is crucial yet challenging. One could use factors such as user historic data evaluation, user specified expertise level on topic, time the user spends on the traversing the trails on general topics of the task, the level of depth they traverse in each sub topic as indicators of expertise level and adjust the recommendations of search trail accordingly.

7. Conclusion

We proposed two evaluation methods that judge search trail-based recommendations in order to improve under-performing users' exploratory search processes. With data from over 300 users, we demonstrated that search trail recommendations could help users at certain times during their information seeking task, thus providing them with more accurate information that meets their task goals (RQ1). We also proved that specifically ordered search trail recommendations facilitate exploratory search tasks more than ad-hoc query suggestions and retrieval features using the goodness of search trail and information coverage metrics we proposed (RQ2).

These research findings have implications for IR evaluation as well as predictions and recommendations for interactive search processes. Specifically, the work reported here provides a new method for evaluating search trails, predicting their outcomes, and using that knowledge to recommend better trails for people working in exploratory search tasks. This research implies that search recommendations should be provided to users as a search trail rather than a simple query suggestion or Web page suggestion allowing the users to navigate the search process in a meaningful way to reach their search goal. For example, providing query auto completion or suggested queries alone might not help specially during exploratory searches, but showing the different paths they could take with each query trail would better assist users to navigate the unknown areas of the topic/task goal they are trying to find. These research findings pave way for future directions in designing and developing better exploratory search guidance platforms that would enable to show search trails using graphs or web traversal paths that indicate which type of paths would lead to better coverage of information on the specific topic or are in concern. Further implications of this research would be to enable exploratory search systems adapting to searcher intention and searcher goals based on the different search trails that they undertake and also enabling the searchers to dynamically visualize the goodness or effectiveness of their search process using metrics that are presented to them. The availability of evaluation metrics would enable searchers to be more effective and be aware of their performance thus making them be more productive than blindly searching for information in vacuum.

Future work would examine traversed information space to determine when gathered information reaches its peak during the search process. Further investigations would identify which search trail characteristics enhance users' information retrieval. In the future, one could design a user study that provides dynamic search trail recommendations and evaluates searchers' reactions using the proposed metrics in real time to improve users' search processes and help them achieve their search task goals.

Acknowledgment

The work reported here is supported by Institute of Museum and Library Services (IMLS) Early Career Development Grant #RE-04-12-0105-12. Authors would also like to thank Prof. Tomasz Imielinski for his suggestions and advice on this research work. We extend our thanks to Dr. Simon Knight and Matthew Mitsui who were instrumental in conducting some of the user studies. Further we would like to acknowledge all the others who helped in conducting the studies and all users who participated in the various user studies explained in this article.

Appendix A. ODP Categorization based Evaluation.

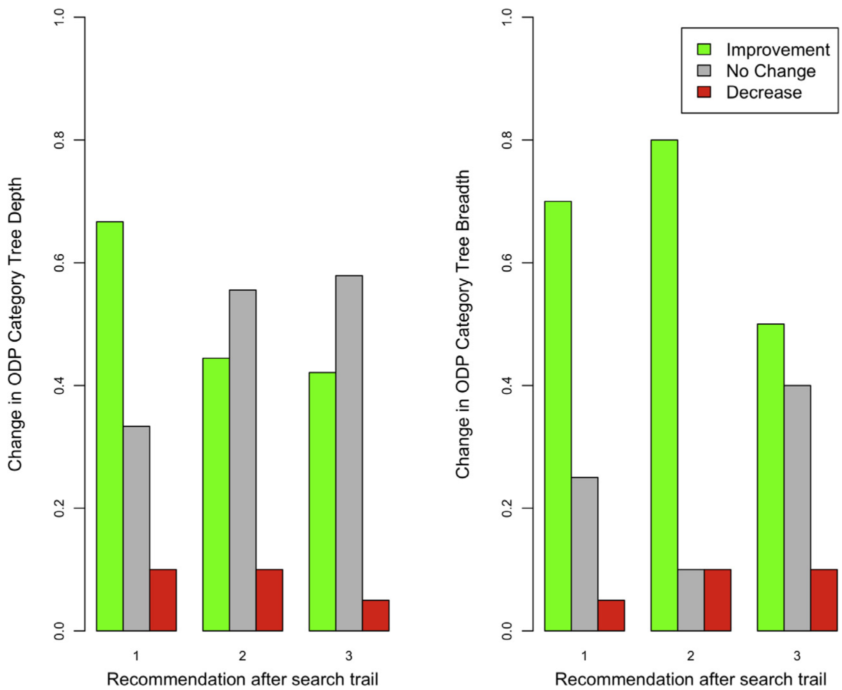

Fig. A.5. Search trail recommendation effect for Global warming task.

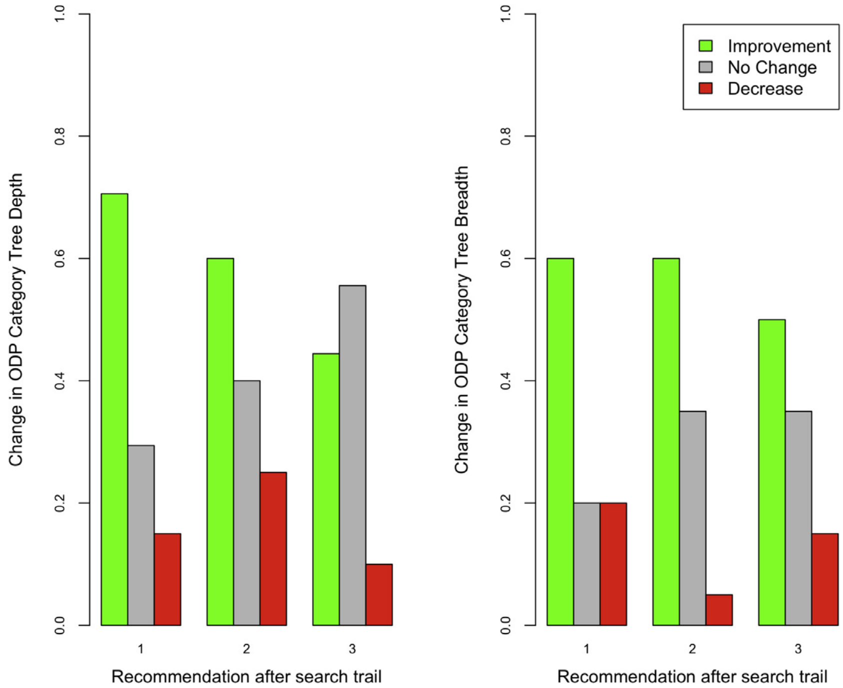

Fig. A.6. Search trail recommendation effect for Red Yeast Rice task.

Appendix B. Likelihood of Discovery based Evaluation.

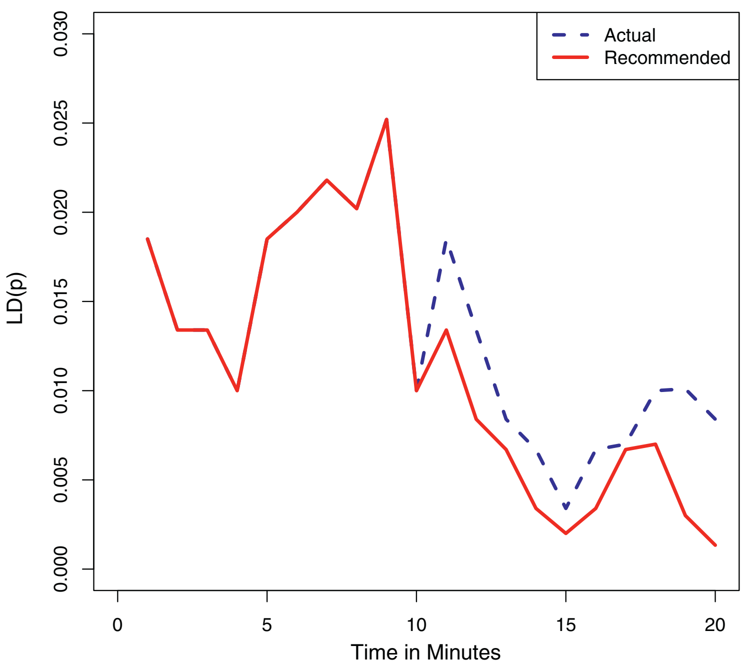

Fig. B.7. LD before and after recommendation at each points in time.

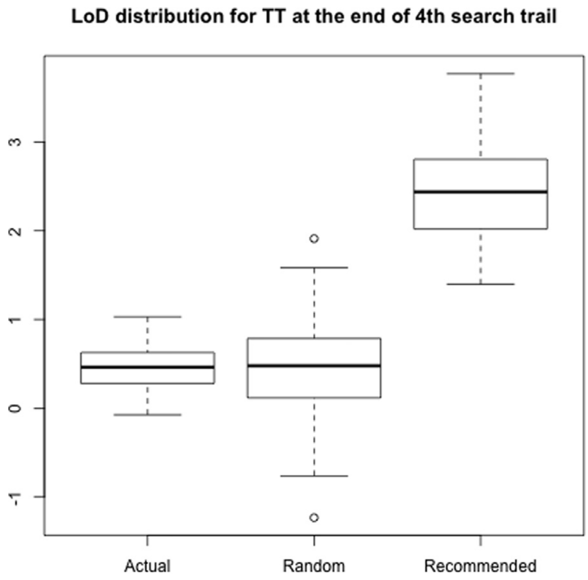

Fig. B.8. *LD* comparison for 4th search trail.

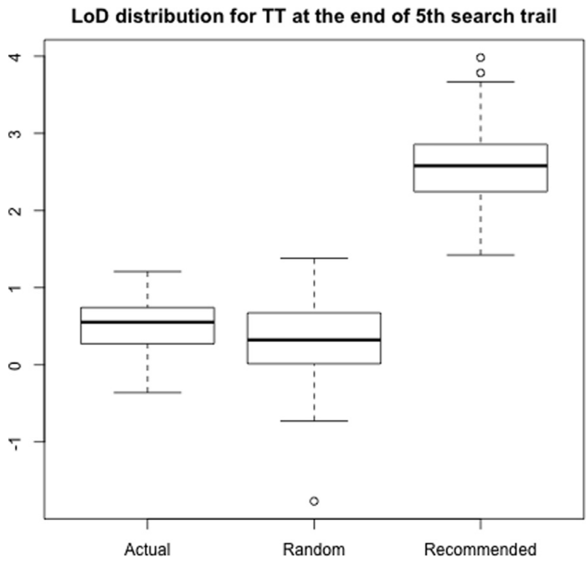

Fig. B.9. *LD* comparison for 5th search trail.

Appendix C. Goodness of search trails based Evaluation.

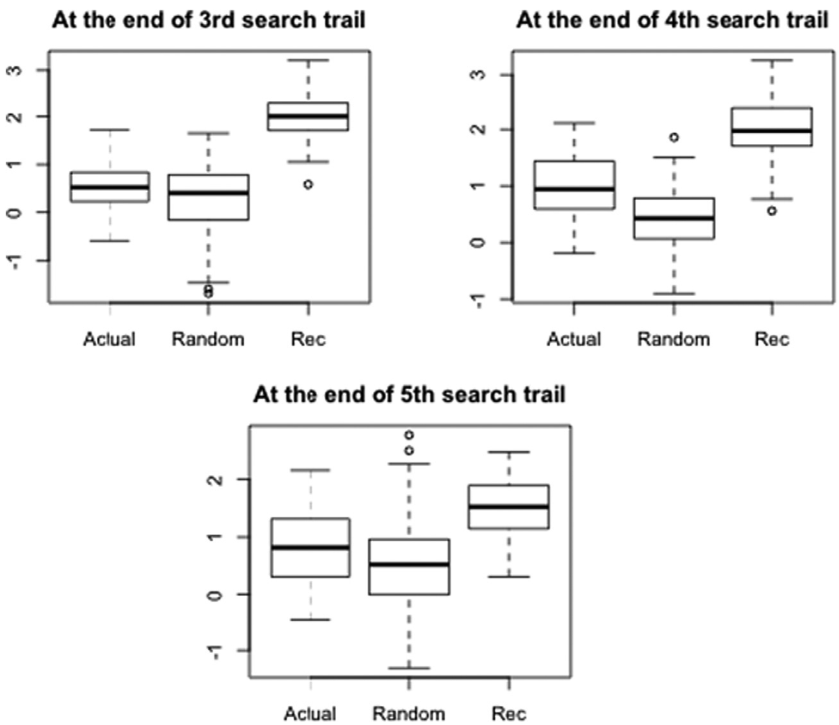

Fig. C.10. *Goodness*(p, p_i) comparison for dataset: Health.

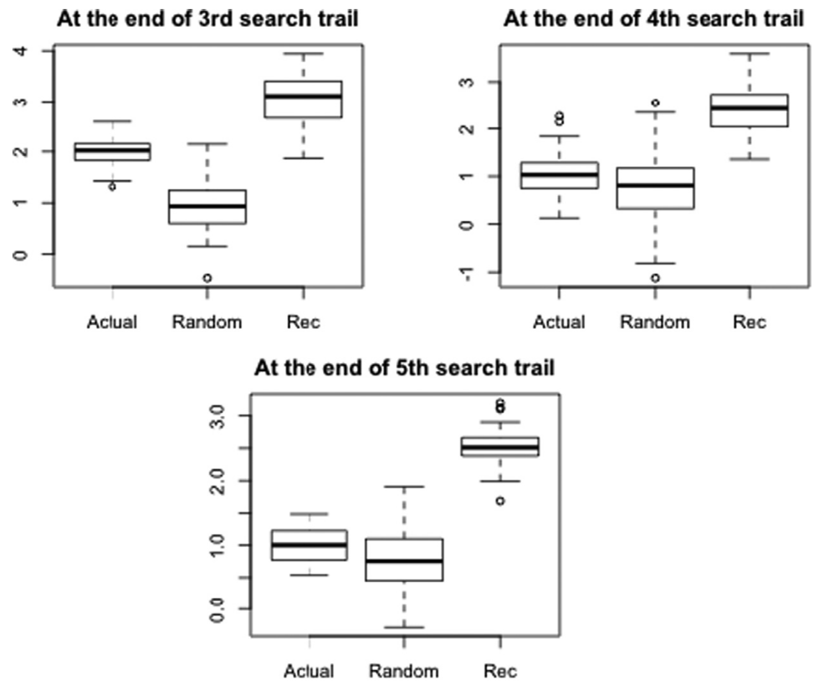

Fig. C.11. *Goodness*(p, p_i) comparison for dataset: Global warming.

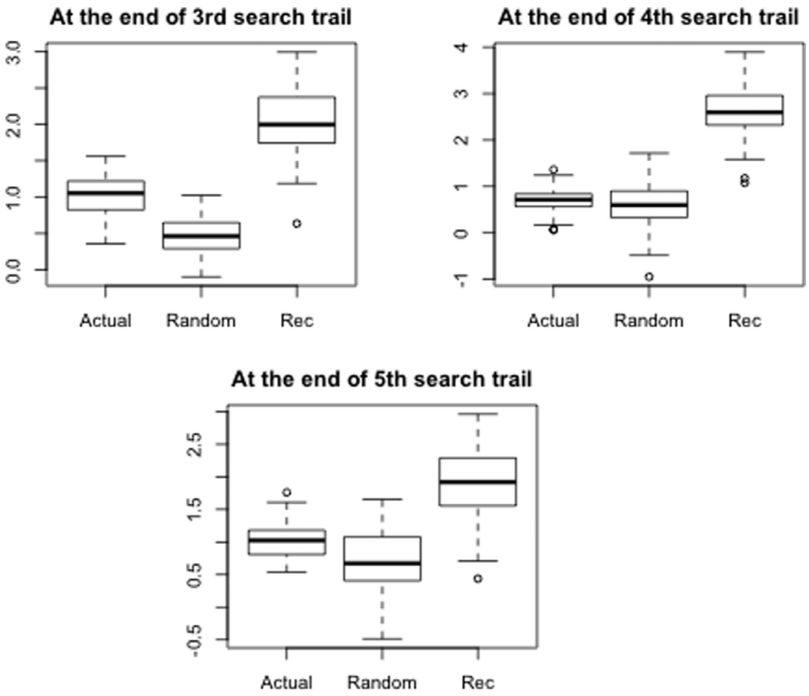

Fig. C.12. *Goodness*(p, p_i) comparison for dataset: Quiet Revolution.

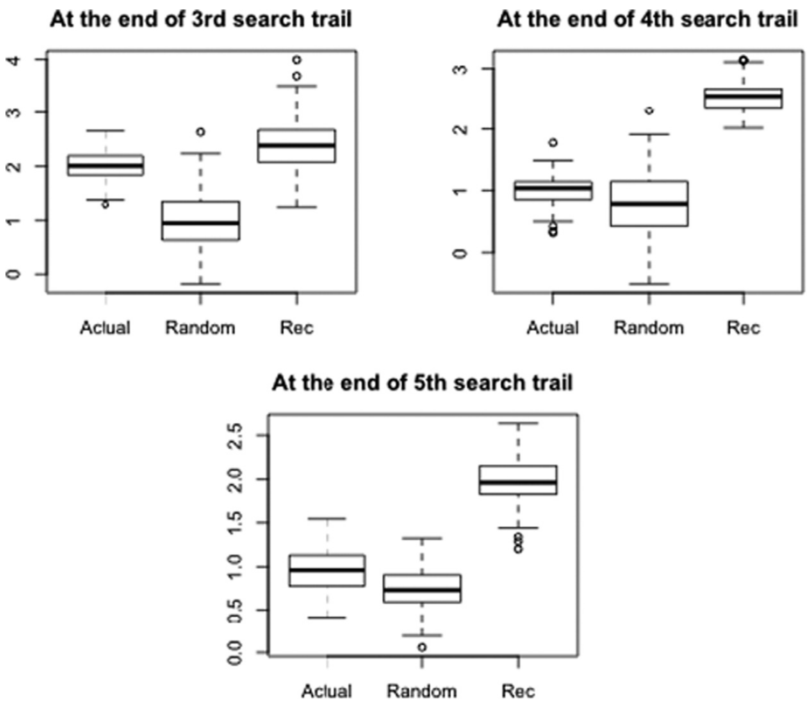

Fig. C.13. *Goodness*(p, p_i) comparison for dataset: Red Yeast Rice.

References

Athukorala, K., Glowacka, D., Jacucci, G., Oulasvirta, A., & Vreeken, J. (2016). Is exploratory search different? a comparison of information search behavior for exploratory and lookup tasks. *Journal of the Association for Information Science and Technology*, *67*(11), 2635–2651. doi:10.1002/asi.23617.

Balog, K., Bron, M., & De Rijke, M. (2011). Query modeling for entity search based on terms, categories, and examples. *ACM Transactionson Information Systems*, *29*(4), 22:1–22:31. doi:10.1145/2037661.2037667.

Basu, C., Hirsh, H., & Cohen, W. (1998). Recommendation as classification: Using social and content-based information in recommendation. In *In proceedings of the fifteenth national conference on artificial intelligence* (pp. 714–720). AAAI Press.

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review, 13*(5), 407–424. doi:10.1108/eb024320.

Bawden, D. (1986). Information systems and the stimulation of creativity. *Journal of Information Science, 12*(5), 203–216. doi:10.1177/016555158601200501.

Belkin, N., Oddy, R., & Brooks, H. (1982). Ask for information retrieval: Part 1. background and theory. *Journal of Documentation, 38*(2), 61–71. doi:10.1108/eb026722.

Borgman, C. L., Hirsh, S. G., & Hiller, J. (1996). Rethinking online monitoring methods for information retrieval systems: From search product to search process. *Journal of the American Society for Information Science, 47*(7), 568–583. doi:10.1002/(SICI)1097-4571(199607)47:7⟨568::AID-ASI8⟩3.0.CO;2-S.

Brando, W. C., Santos, R. L. T., Ziviani, N., de Moura, E. S., & da Silva, A. S. (2014). Learning to expand queries using entities. *Journal of the Association for Information Science and Technology, 65*(9), 1870–1883. doi:10.1002/asi.23084.

Budd, J. M. (2004). Relevance: Language, semantics, philosophy. *Library Trends, 52*, 447–462.

Cai, F., & de Rijke, M. (2016). Learning from homologous queries and semantically related terms for query auto completion. *Information Processing & Management, 52*(4), 628–643. http://dx.doi.org/10.1016/j.ipm.2015.12.008.

Cai, F., Wang, S., & de Rijke, M. (2016). Behavior-based personalization in web search. *Journal of the Association for Information Science and Technology*, n/a–n/a. doi:10.1002/asi.23735.

Downey, D., Dumais, S., & Horvitz, E. (2007). Models of searching and browsing: Languages, studies, and applications. In *Proceedings of the 20th international joint conference on artifical intelligence*. In *IJCAI'07* (pp. 2740–2747). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Egusa, Y., Saito, H., Takaku, M., Terai, H., Miwa, M., & Kando, N. (2010). Using a concept map to evaluate exploratory search. In *Proceedings of the third symposium on information interaction in context*. In *IIiX '10* (pp. 175–184). New York, NY, USA: ACM. doi:10.1145/1840784.1840810.

Foster, A., & Ford, N. (2003). Serendipity and information seeking: An empirical study. *Journal of Documentation, 59*(3), 321–340. doi:10.1108/00220410310472518.

Fox, S., Karnawat, K., Mydland, M., Dumais, S., & White, T. (2005). Evaluating implicit measures to improve web search. *ACM Transactions on Information Systems, 23*(2), 147–168. doi:10.1145/1059981.1059982.

Hassan, A., & White, R. W. (2013). Personalized models of search satisfaction. In *Proceedings of the 22nd ACM international conference on conference on information and knowledge management*. In *CIKM '13* (pp. 2009–2018). New York, NY, USA: ACM. doi:10.1145/2505515.2505681.

Hassan Awadallah, A., White, R. W., Pantel, P., Dumais, S. T., & Wang, Y.-M. (2014). Supporting complex search tasks. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*. In *CIKM '14* (pp. 829–838). New York, NY, USA: ACM. doi:10.1145/2661829.2661912.

He, J., Qvarfordt, P., Halvey, M., & Golovchinsky, G. (2016). Beyond actions: Exploring the discovery of tactics from user logs. *Information Processing & Management, 52*(6), 1200–1226. http://dx.doi.org/10.1016/j.ipm.2016.05.007.

He, Q., Jiang, D., Liao, Z., Hoi, S. C. H., Chang, K., Lim, E.-P., & Li, H. (2009). Web query recommendation via sequential query prediction. In *Proceedings of the 2009 IEEE international conference on data engineering*. In *ICDE '09* (pp. 1443–1454). Washington, DC, USA: IEEE Computer Society. doi:10.1109/ICDE.2009.71.

Hendahewa, C., & Shah, C. (2015). Implicit search feature based approach to assist users in exploratory search tasks. *Information Processing & Management, 51*(5), 643–661. http://dx.doi.org/10.1016/j.ipm.2015.06.004.

Karisani, P., Rahgozar, M., & Oroumchian, F. (2016). A query term re-weighting approach using document similarity. *Information Processing & Management, 52*(3), 478–489. http://dx.doi.org/10.1016/j.ipm.2015.09.002.

Kuhlthau, C. (2004). Seeking meaning: A Process approach to library and information services. *Information management, policy, and services*. Libraries Unlimited.

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science, 42*(5), 361–371 10.1002/(SICI)1097-4571(199106)42:5-361::AID-ASI6>3.0.CO;2-#.

Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM, 49*(4), 41–46. doi:10.1145/1121949.1121979.

Marchionini, G., & Shneiderman, B. (1988). Finding facts vs. browsing knowledge in hypertext systems. *Computer, 21*(1), 70–80. doi:10.1109/2.222119.

Miranda, T., Claypool, M., Claypool, M., Gokhale, A., Gokhale, A., Mir, T., … Sartin, M. (1999). Combining content-based and collaborative filters in an online newspaper. *In proceedings of ACM SIGIR workshop on recommender systems*.

Niu, X., & Kelly, D. (2014). The use of query suggestions during information search. *Information Processing & Management, 50*(1), 218–234. http://dx.doi.org/10.1016/j.ipm.2013.09.002.

Olston, C., & Chi, E. H. (2003). Scenttrails: Integrating browsing and searching on the web. *ACM Transactions on Computer-Human Interaction, 10*(3), 177–197. doi:10.1145/937549.937550.

Park, D. H., Kim, H. K., Choi, I. Y., & Kim, J. K. (2012). A literature review and classification of recommender systems research. *Expert Systems with Applications, 39*(11), 10059–10072. http://dx.doi.org/10.1016/j.eswa.2012.02.038.

Pirolli, P., & Card, S. K. (1999). Information foraging. *Psychological Review, 106*, 643–675.

Pirolli, P., & Fu, W. T. (2003). Snif-act: A model of information foraging on the world wide web. In *Proceedings of the ninth international conference on user modeling* (pp. 45–54). Springer-Verlag.

Qu, Y., & Furnas, G. W. (2008). Model-driven formative evaluation of exploratory search: A study under a sensemaking framework. *Information Processing & Management, 44*(2), 534–555. http://dx.doi.org/10.1016/j.ipm.2007.09.006. Evaluating Exploratory Search SystemsDigital Libraries in the Context of Users Broader Activities.

Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (2010). *Recommender systems handbook* (1st). New York, NY, USA: Springer-Verlag New York, Inc.

Saari, D. G. (2000). Mathematical structure of voting paradoxes: Ii. positional voting. *Economic Theory, 15*(1), 55–102.

Shah, C., Liu, J., González-Ibáñez, R., & Belkin, N. (2012). Exploration of dynamic query suggestions and dynamic search results for their effects on search behaviors. *Proceedings of the American Society for Information Science and Technology, 49*(1), 1–10. doi:10.1002/meet.14504901135.

Singla, A., White, R., & Huang, J. (2010). Studying trailfinding algorithms for enhanced web search. In *Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval*. In *SIGIR '10* (pp. 443–450). New York, NY, USA: ACM. doi:10.1145/1835449.1835524.

Smith, C. L., Gwizdka, J., & Feild, H. (2016). Exploring the use of query auto completion: Search behavior and query entry profiles. In *Proceedings of the 2016 ACM on conference on human information interaction and retrieval*. In *CHIIR '16* (pp. 101–110). New York, NY, USA: ACM. doi:10.1145/2854946.2854975.

Vidinli, I. B., & Ozcan, R. (2016). New query suggestion framework and algorithms: A case study for an educational search engine. *Information Processing & Management, 52*(5), 733–752. http://dx.doi.org/10.1016/j.ipm.2016.02.001.

Wexelblat, A., & Maes, P. (1999). Footprints: History-rich tools for information foraging. In *Proceedings of the SIGCHI conference on human factors in computing systems*. In *CHI '99* (pp. 270–277). New York, NY, USA: ACM. doi:10.1145/302979.303060.

White, R. W., Bilenko, M., & Cucerzan, S. (2007). Studying the use of popular destinations to enhance web search interaction. In *Proceedings of the 30th annual international ACM SIGIR conference on research and development in information retrieval*. In *SIGIR '07* (pp. 159–166). New York, NY, USA: ACM. doi:10.1145/1277741.1277771.

White, R. W., & Huang, J. (2010). Assessing the scenic route: Measuring the value of search trails in web logs. In *Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval*. In *SIGIR '10* (pp. 587–594). New York, NY, USA: ACM. doi:10.1145/1835449.1835548.

White, R. W., & Roth, R. A. (2009). Exploratory search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services, 1*(1), 1–98.