Data and Information Management, 2021; 5(1): 228–241

2020 ASIS&T Asia-Pacific Regional Conference (Virtual Conference), December 12-13, 2020, Wuhan, China

Open Access

Afeng Wang, Yiming Zhao*, Yijin Chen

Information Search Trail Recommendation Based on Markov Chain Model and Case-based Reasoning

https://doi.org/10.2478/dim-2020-0047
received August 15, 2020; accepted September 15, 2020.

Abstract: An information search trail recommendation method based on the Markov chain model and case-based reasoning is proposed. A laboratory user experiment was designed to evaluate the proposed method. The experimental results demonstrated that novice searchers have a positive attitude toward the search trail recommendation and a willingness to use the recommendation. Importantly, this study found that the search trail recommendation could effectively improve novice searchers' search performance. This finding is mainly reflected in the diversity of information sources and the integrity of the information content of the search results. The proposed search trail recommendation method extends the application scope of information recommendations and provides insights to improve the organization and management of online information resources.

Keywords: Web search, search process, search trail recommendation, Markov chain, case-based reasoning

1 Introduction

Using Web search engines, people can obtain a piece of information or navigate to a target website quickly. However, frequently, many information search tasks are complex or exploratory, such as acquiring knowledge items of a particular subject or writing a course thesis. The information needs of these tasks include multiple aspects or steps, and the search processes often require access to different types of websites. However, in practice, many novice searchers find that locating the appropriate websites efficiently is difficult. As a result, many specialist websites fail to establish connections with target users. Fortunately, expert searchers with certain domain knowledge, search experience, or skills can conduct efficient searches and locate the appropriate websites or website sequences to find what they are looking for (Tabatabai & Shore, 2005). White, Dumais, and Teevan (2009) recommend that query suggestions and website recommendations generated through domain experts' search history could be provided to novice searchers to help them gain expertise. They even envisaged developing such Web search support services in future research.

Some previous studies have attempted to support new searchers by providing webpage recommendations (Hendahewa & Shah, 2017; White, Bilenko, & Cucerzan, 2007), optimizing webpage rankings (Ziegler, McNee, Konstan, & Lausen, 2005), and query expansion (Smith, Gwizdka, & Field, 2016; Huang, Wang, Zhang, & Liu, 2020). However, such supports are often insufficient to meet complex information needs. Novice searchers may need support that alerts them to the steps, or a webpage sequence, or websites required for complex task completion.

Previous studies have shown that trails or tours consisting of filtered documents or webpages can reveal the value of user search processes, and trail-based search recommendations could improve new searchers' overall search performance. For example, White and Huang (2010) demonstrated that following search trails provides significant additional benefits to searchers in terms of coverage, diversity, novelty, and utility over origin pages

***Corresponding author: Yiming Zhao,** School of Information Management, Wuhan University, Wuhan, China,
Email: zhaoyiming@whu.edu.cn
Afeng Wang, School of Information Management, Wuhan University, Wuhan, China
Yijin Chen, School of Economic and Management, South China Normal University, Guangzhou, China

and destination pages. Hendahewa and Shah (2017) found that, with the help of search trail recommendations, searchers can find more information across multiple facets and dig deeper into the detail associated with certain facets.

However, most search trails mined from logs often only contain pages visited after a single query rather than all pages or websites visited during the full search process of a search task. Besides, to date, proposed methods for finding trails often focus on trails consisting of webpages that may be sensitive to Web dynamism (e.g., dead links or changing content), rather than a more general level of abstraction (e.g., website categories) that may be more widely applicable. Moreover, showing the trails to searchers directly on the search engine result page (SERP) is also an unaddressed challenge (Hassan & White, 2012). Addressing the shortcomings of previous studies, our primary goal is to present and evaluate a method to create search trails that can help novice searchers perform complex search tasks. We focus on (i) how to model and recommend search-task relevant trails of expert searchers and (ii) whether the search trail recommendation is useful for novice searchers.

The remainder of this paper is organized as follows. An extensive literature review of search trail recommendation studies, the utility of the Markov chain model for Web path analysis, and the case-based reasoning (CBR) approach are presented in Section 2. In Section 3, we introduce our proposed information search trail recommendation method based on the Markov chain model and the CBR approach. In Section 4, we describe a user study designed to evaluate the search trail recommendation method. The user study was conducted in a laboratory environment, and the results were analyzed to determine how novice searchers evaluate the effectiveness of search trail recommendations. The implications and limitations of this research and suggestions for future research are summarized in Section 5.

2 Related Work

Several research areas are relevant to the current study: (i) search trail modeling and recommendations, (ii) the Markov chain model, and (iii) CBR searching. In this section, we describe relevant studies in each area in more detail and discuss how the current research could extend them.

2.1 Search Trails Modeling and Recommendation

Some researchers defined the *search trail* as a search path that comprises query and post-query pages visited to carry out relevant studies (Bilenko & White, 2008; Singla, White, & Huang, 2010; Hendahewa & Shah, 2017). Recently, Capra and Arguello (2019) defined a *search trail* as an interactive visualization of how a searcher performs a search task. The visualization may include queries posted, sites or pages visited, and annotations made. It is evident that the latter definition has a larger scope and greater applicability. Thus, we adopted the second definition to carry out our research.

Berrypicking, orienteering, and information foraging are three well-known information seeking models related to search trails. The berrypicking model describes the movement between information sources associated with dynamic information needs (Bates, 1989). The orienteering analogy was proposed to understand searchers' information seeking strategies (O'Day & Jeffries, 1993). Information foraging emphasizes how information searchers use clues left by previous visitors to find information (Pirolli & Card, 1999).

Interaction logging has implicitly made us all trailblazers of search trails (White & Huang, 2010). Trails or tours created by previous searchers form links between stored information resources that can help other searchers make better decisions about information source selection during the search process. Trigg (1988) proposed a guided tour consisting of a sequence of hypertext pages to alleviate disorientation for new searchers. Wheeldon and Levene (2003) proposed an algorithm to generate trails as trees to assist searchers in Web navigation. The results showed that participants found these trails useful for navigation. Singla, White, and Huang (2010) proposed *trailfinding* methods to support Web searches by identifying query-relevant trails from logs that could be shown to complement or replace traditional search result lists. Recently, Hendahewa and Shah (2017) demonstrated that recommending search trails of each query to struggling users in exploratory search tasks could better assist them to find the information they were seeking. Moreover, the results showed that the order of the recommended search trails plays an important role. Capra and Arguello (2019) discovered that task determinability is a significant factor that affects whether to recommend search trails. Besides, they found that the system should provide trails with the same scope as the searcher's task. Search trail recommendation systems, such as *WebWatcher* (Joachims, Freitag, & Mitchell, 1997), *Footprints* (Wexelblat & Maes, 1999), *ScentTrails* (Olston & Chi, 2003), *Volant* (Pandit & Olston, 2007), and *SearchGuide* (Capra, & Arguello, 2019), highlight candidate pages to indicate paths to search results. Research into these systems has shown that they can effectively improve user search performance.

It is worth noting that, in terms of search quality, Yuan and White (2012) found that better quality search trails are created by domain experts. Therefore, it seems wise that experts construct guided search trails or tours. Thus, we considered this when we constructed the search trails recommendation database.

The approach we describe in this article extends existing studies in several different ways. First, we organize potential search-task relevant trail recommendations by mining expert searchers' search logs using the Markov chain model and CBR approach. Second, search trails are constructed at the website category level rather than at the webpage or document level. This approach allows us to detect higher-level patterns of search behavior. Third, we aim to provide searchers with a holistic view of the search trail and one-step recommendations during the search process, which may lead to more effective information search strategies for novice searchers.

2.2 Markov Chain Model

In a search trail recommendation system, modeling the search trail is a significant component; therefore, the next step is to select an appropriate mathematical model to model experts' search trails. Probabilistic models have been applied successfully to many time series prediction problems. In particular, Markov chains and Markov models have achieved great success in sequence generation.

Markov chains allow the system to dynamically model the URL access patterns observed in the logs based on the previous state. In addition, the Markov chain model can be used in generative models to obtain trails automatically. The Markov state transition matrix can be viewed as a "user traversal" representation of the Web space (Sarukkai, 2000).

The utility of the Markov chain model has been demonstrated in many domains, such as link prediction and path analysis (Sarukkai, 2000), the personalized recommendation in information retrieval systems (Liu, Huang, & An, 2007), exhibition booth visit recommendations (Moon, Kim, & Ryu, 2013), and path prediction in Internet of things (IoT) systems (Piccialli, Cuomo, Giampaolo, Casolla, & di Cola, 2020). For example, Sarukkai (2000) presented an algorithm for search tour generation using Markov chains and demonstrated that Markov chains are useful tools for Web link sequence modeling and path analysis. Liu, Huang, and An (2007) proposed a mixture of Markov models to cluster searchers, capture the sequential relationships in searchers' access histories, and provide searchers with personalized recommendations.

In this paper, we apply the Markov chain model to analyze the search trail of site nodes. The state transition matrix of the Markov chain model can be considered as a "weighted traversal" representation of the user's model of the Web space, and further analysis can be performed on this matrix, such as link relationships between different categories of websites.

2.3 CBR Searching

The principle of the CBR approach is analogical reasoning, and its basic idea is that new problems can be solved with the help of the solutions to similar past problems (Gentner, 1983; Hüllermeier, 2007). The knowledge base of the CBR system consists of a collection of cases and a set of search criteria used to retrieve cases similar to the target problem (Althuizen & Wierenga, 2014). A historical case in the case base is represented as c = (*Specification*, *Solution*), where *Specification* is the description of the problem consisting of n features, and *Solution* provides the solution to the problem. The CBR approach was originally applied in the field of artificial intelligence (Aamodt & Plaza, 1994). And now, the CBR approach has since been applied in many other fields, such as business (Gavetti, & Rivkin, 2005; Goldstein, 2001; Gregan-Paxton & Cote, 2000), medical diagnosis (Bichindaritz & Marling, 2006), information seeking (He, Erdelez, Wang, & Shyu, 2008; Alptekin & Büyüközkan, 2011), engineering (Shokouhi, Skalle, & Aamodt, 2014), architecture, and law (Bridge, Göker, McGinty, & Smyth, 2006; Hüllermeier, 2007). For example, doctors may benefit from using a CBR system that accesses the case of a previously treated patient with symptoms similar to those of a new patient (Bichindaritz & Marling, 2006).

Although many studies have focused on search trail recommendations, few studies consider CBR searching. Moore, Erdelez, and He (2006) conducted a controlled experiment that demonstrated the difference between traditional keyword searching and CBR searching. However, the authors did not explain why the difference occurs. He and Tian (2017) conducted an 8-year longitudinal analysis of the query logs of a Web-based case library system. They found those return users employed CBR searching much more frequently than one-time users.

In addition, return users tended to use more query terms to find information.

We assume that a recommendation based on the CBR approach might recommend search strategies that reflect how experts conducted search tasks (e.g., search trails at the websites level) to novice searchers who perform the same or similar tasks, which will extend the scope of search recommendation. Therefore, in this study, the CBR method was adopted by storing expert searchers' search trails of site nodes in the case base and generating recommendations for novice searchers who need to solve the same or a similar task.

3 Search Trail Recommendation Method Based on Markov Chain Model and CBR Approach

In this section, we describe the search trail recommendation procedure. First, the Markov chain model was used to extract the search trails of the user information search process. The full search trail, that is, the sequence of Internet information sources that experts visited, and one-step trail, that is, the transfer probability of Internet information sources categories, are identified and stored in the form of a case base that serves as the data source of information search trail recommendations for novices. Then, based on the CBR approach, the expert search trail is re-used as recommended content to improve a novice's information search experience and information resource utilization efficiency.

3.1 Search-Task Relevant Search Trails Generation

The search task is the original driving force of a searcher's information seeking behavior. The search task shapes the searcher's interaction with various information sources during the search process (Li & Belkin, 2010). Task attributes, such as type, complexity, and goal, greatly influence how a searcher formulates their search strategy and their search behavior (Kim, 2009; Li, 2009; Li & Belkin, 2010).

First, the sequence in which searchers access Internet sources during a search task must be identified. A sequence of Internet sources accessed by a searcher when performing a search task is described in Equation (1) as follows:

$$\boldsymbol{SrcSeq} = \left(Src_1, Src_2, \ldots, Src_i, Src_j, \ldots, Src_l\right) \quad (1)$$

To handle the large volume of Internet sources, we propose using Internet source categories to reduce the state space of Markov models as follows:

$$\begin{aligned}\boldsymbol{SrcCatSeq} = \big(&SrcCat_1, SrcCat_2, \ldots,\\ &SrcCat_i, SrcCat_j, \ldots, SrcCat_l\big)\end{aligned} \quad (2)$$

In Equation (2), $SrcCat_i \in Cat = \{cat_1, cat_2, \ldots, cat_n\}$, n represents the number of Internet sources categories and indicates that there are n states in the Markov chain model.

Then, deploying the CBR approach, the Markov chain model of search-task relevant search trails can be expressed as follows:

$$\boldsymbol{tmc} = \left(\boldsymbol{T}, Cat, S, SC, \boldsymbol{A}\right) \quad (3)$$

Here, $\boldsymbol{T}$ is a set of properties that describe the search task and also represents a set of criteria for retrieving cases that are similar to the target search task. S denotes collections of Internet source sequences and SC represents information source category sequences generated by expert searchers when performing certain information search tasks. $\boldsymbol{A}$ is the transition probability matrix of the search trail Markov model of certain information search tasks. The transition probability matrix can be trained using historical data. Without loss of generality, this study uses the principle of maximum likelihood to estimate $\boldsymbol{A}$. Mathematical representations of $\boldsymbol{T}$, S, SC, and $\boldsymbol{A}$ are given as follows:

$$\boldsymbol{T} = \left(a_1, \ldots, a_m\right) \quad (4)$$

$$S = \left\{\boldsymbol{SrcSeq}_1, \boldsymbol{SrcSeq}_2, \ldots, \boldsymbol{SrcSeq}_k\right\} \quad (5)$$

$$SC = \left\{\boldsymbol{SrcCatSeq}_1, \boldsymbol{SrcCatSeq}_2, \ldots, \boldsymbol{SrcCatSeq}_k\right\} \quad (6)$$

$$\boldsymbol{A} = \left(p_{ij}\right) = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1j} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2j} & \cdots & p_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{i1} & p_{i2} & \cdots & p_{ij} & \cdots & p_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nj} & \cdots & p_{nn} \end{bmatrix} \quad (7)$$

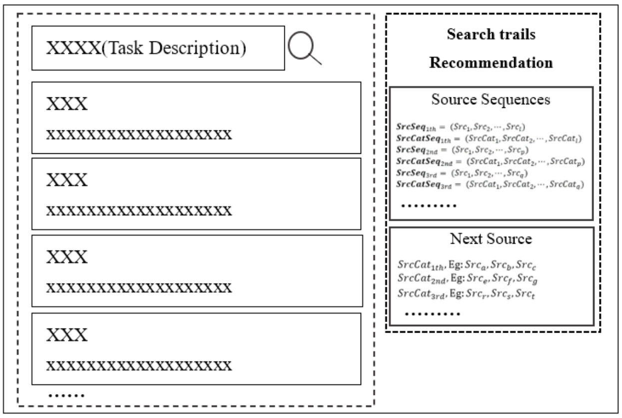

Figure 1. Search trail recommendation system initial interface.

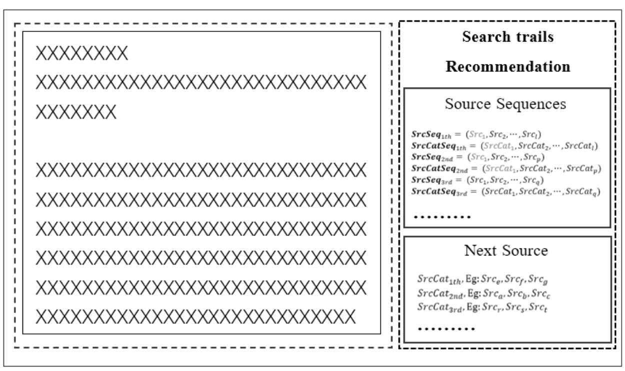

Figure 2. Search trail recommendation for a specific Internet source.

The element in $\boldsymbol{A}$ is described in Equation (8):

$$p_{ij} = \frac{C(cat_i, cat_j)}{\sum_{j=1}^{n} C(cat_i, cat_j)} \tag{8}$$

In Equation (8), $C(cat_j, cat_i)$ is the count of the number of times cat_j follows cat_i in the dataset SC. Although Markov chains have traditionally been used to characterize the asymptotic properties of random variables, we use the transition matrix to estimate short-term link predictions of Internet sources. An element of the matrix $\boldsymbol{A}$, say p_{ij} can be interpreted as the probability of transitioning from state cat_i to cat_j in a single step. Similarly, an element of $\boldsymbol{A} * \boldsymbol{A}$ will denote the probability of transitioning from one state to another in two steps, and so on. To provide novices with a recommendation for the next possible Internet source to be selected, we feed the current Internet source into the model and let the transition matrix determine the optimal candidates.

Then, based on the Markov chain model, the search-task relevant search trails case base, which we refer to as *TMC*, is formally expressed as follows.

$$TMC = \left\{ \boldsymbol{tmc}_i \mid \boldsymbol{tmc}_i = \left(\boldsymbol{T}_{tmc_i}, Cat_{tmc_i}, S_{tmc_i}, SC_{tmc_i}, \boldsymbol{A}_{tmc_i} \right), i = 1, \dots, N \right\} \tag{9}$$

In Equation (9), $\boldsymbol{tmc}_i$ represents the search trail recommendation case of a certain type of search task, and N represents the number of cases in the case base.

3.2 Search Trail Recommendation Procedure

3.2.1 Retrieving Similar Relevant Cases

To retrieve the expert search trails for the recommendation, the similarity of task characteristic attributes between the current search task (***SrchTsk***) and the task in the search trails case base should be calculated. The greater the similarity value, the more relevant the case. When the similarity value is higher than a certain threshold, the case will be added to the expert search trail recommendation case set, *ET*, which is expressed as follows:

$$\begin{gathered} ET\left(\boldsymbol{SrchTsk}\right) = \\ \left\{ \boldsymbol{tmc}_i \mid sim\left(\boldsymbol{SrchTsk}, \boldsymbol{T}_{tmc_i}\right) \geq q, \forall \boldsymbol{tmc}_i \in TMC \right\} \end{gathered} \tag{10}$$

3.2.2 Search Trail Recommendation During Search Process

The initial interface of the search trail recommendation system is shown in Figure 1. After the searcher inputs the description of the search task, the search trail recommendations list ranked by similarity will be generated in a timely manner. The left part of the interface presents the default search engine's results list, and the right part presents the search trail recommendations. The upper half of the right part presents the *S* and *SC* lists, that is, the full search trail list. The lower half displays the Internet source categories that expert searchers are most likely to investigate next under the current Internet source, that is, the one-step trail recommendations.

If the vector $\boldsymbol{s}(t+1)$ denotes the probability vector for all states at time $t+1$. Given the history of "Internet information sources" of the novice searcher $\boldsymbol{s}(t-k), \boldsymbol{s}(t-k+1), \dots, \boldsymbol{s}(t)$, each Internet information source could be represented as a vector with probability 1 at that state for that time (denoted $\boldsymbol{i}(t-k), \boldsymbol{i}(t-k+1), \ldots, \boldsymbol{i}(t)$). The Markov chain model estimation of the probability of being in a given state at time $t+1$ is shown in Equation (11).

$$\hat{\boldsymbol{s}}(t+1) = \hat{\boldsymbol{i}}(t)\boldsymbol{A} \tag{11}$$

Then, the one-step search trail recommendations list ranked in descending probability value in vector $\boldsymbol{s}(t+1)$ would be provided. The interface for a specific Internet source is represented in Figure 2. The Internet source cases under each category are given, as shown in Figures 1 and 2.

4 Evaluation

In this section, we describe a user study that was designed to evaluate the effectiveness of the search trail recommendation method proposed in this paper. First, the search trails generation method proposed in Section 3 is used to identify and organize the search trails generated by expert searchers when performing some search tasks. Then, novice searchers are organized to perform the same or similar tasks, and the relevant search trail recommendations are provided during the search process. Finally, based on novice searchers' attitudes (e.g., perceived usefulness, satisfaction, and acceptance) toward the search trail recommendations and their search results performance, the recommendation method proposed in this study can be evaluated. In addition, the characteristics of tasks and searchers that may influence a searcher's willingness to engage with search trails, and their ability to benefit from these search trails could be studied deeply. If the results of the user study show that novice searchers benefit from search trail recommendations, likely post-task trails could be considered in real search system design and even as units of retrieval in practice.

4.1 User Experiment Design

4.1.1 Data Collection

Questionnaire surveys, screen recordings, think-aloud guidelines, and interviews were used to collect data.

Before the experiment, participants were asked to complete an entry questionnaire that was used to collect user demographic data, including search experience and habits, as well as their attitude toward sharing search trails or accepting the search trail recommendations. A software tool recorded all participant activities during the experiment. A think-aloud guideline was designed to elicit participants' cognitive activities during their search performances. After the search experiment, a post-search questionnaire queried the novice searchers' perceptions of their search results, search performance, difficulties, and the search trail recommendations provided during the search process. In addition, we conducted follow-up interviews to explore the novice searchers' comments and suggestions about the search trail recommendations.

4.1.2 Participants

In this experiment, participants were recruited from an iSchool at a national major university in China. Twelve participants, all of whom had experience searching on the Internet, were recruited. Each participant was paid RMB 50 Yuan as compensation.

One senior undergraduate and four postgraduates were recruited as expert searchers (one male, four females, denoted $expert[i], i=1,\ldots,5$). All expert searchers had considerable information literacy training before the experiment, and all were willing to share search trail experience with others. Seven freshmen were recruited as novice searchers (two males, five females, denoted $novice[j], j=1,\ldots,7$). None of the novice searchers had information literacy training experience prior to the experiment.

4.1.3 Materials and Apparatus

4.1.3.1 Tasks

During the experiment, the subjects were asked to solve two information problems: one low complexity task and one high complexity task, as shown in Table 1. In this experiment, the task context refers to an environment that involves a combination of various factors and conditions, such as task product and search time. Participants could search for information on the Internet without any restrictions while performing these search tasks. Each task requires the submission of a document containing the required information collected and organized by the searcher.

4.1.3.2 Setting and Equipment

The information search experiment was conducted in a laboratory. EV software was installed on the computers and the entire experiment was recorded. Both screen

Table 1
Tasks

Type	Scenario and instructions
Class presentation (Low task complexity)	Imagine you are going to share knowledge about "cloud computing" with your classmates in the next class. Please search the Internet for relevant information.
Course thesis (High task complexity)	Imagine you are doing your course thesis, which is a study about artificial intelligence in education/healthcare (choose one of the two topics). Please search the Internet for relevant information.

Table 2
Internet Information Source Categories

Categories		Examples
A	Comprehensive search engine	baidu.com, cn.bing.com
B	Academic search engine	xueshu.baidu.com, academic.microsoft.com
C	E-journal database	cnki.net, jstor.org, link.springer.com
D	Electronic library	sslibrary.com, brill.com, books.google.com
E	Document sharing platform	wenku.baidu.com, doc88.com
F	Online encyclopedia	wiki.tw.lvfukeji.com, baike.baidu.com
G	Online community and social Q&A platform	researchgate.net, bbs.pinggu.org, zhihu.com, zhidao.baidu.com
H	Blog platform	blog.csdn.net, blog.sciencenet.cn
I	Other websites	Such as news websites and government websites

capture and audio recording modes were used to record the entire task performance process and the participants' verbalized thoughts. The questionnaire survey was launched using wjx.cn, a Web-based survey tool.

4.1.4 Measurements

4.1.4.1 Internet Information Source Categories

In this experiment, considering the tasks, we classified Internet information sources into the categories listed in Table 2.

Table 3
Attitude Measurements

Dimension	Items
Perceived usefulness (PU)	The search trail recommendation can improve my search performance.
	The search trail recommendation can improve my search efficiency.
	The search trail recommendation is useful for my information search.
Satisfaction (SA)	I think the decision to use the search trail recommendation is wise.
	I feel enjoyable when I search with the search trail recommendation.
	I am satisfied with the search trail recommendation.
Acceptance (AC)	I am willing to use the search trail recommendation.
	I intend to use the search trail recommendation regularly in the future.
	I intend to use the search trail recommendation more often in the future.

4.1.4.2 Attitude Measurements

This study adopted and modified some attitude measurements (Bhattacherjee, 2001; Lin & Wang, 2012) to investigate novice searchers' perceived usefulness, satisfaction, and acceptance attitudes toward the search trail recommendations provided in this experiment. The measurement items are listed in Table 3. Each evaluation dimension score is the average of all measurement items, where 1 was the lowest score and 7 was the highest score.

4.1.5 Procedure

Prior to participating in the experiment, the participants were given a description of the experiment and completed the entry questionnaire. After reading the search task assignment, they were asked to read the "think-aloud guideline" carefully. In addition, participants labeled as novice searchers received a list of recommended search trails generated by participants labeled as expert searchers performing similar tasks, and they were asked to use these recommendations as much as possible. After finishing the search, all participants were asked to complete a post-search questionnaire. Besides, we conducted an exit interview with each participant.

Table 4
Source Categories Transition Probability Matrix of Class Presentation Task

→	A	B	C	D	E	F	G	H	I	Σ
A	7%				7%	29%	21%	7%	29%	100%
B			50%						50%	100%
C	33%					33%			33%	100%
D										100%
E	50%		50%							100%
F	50%						25%		25%	100%
G		17%	17%			17%			50%	100%
H					100%					100%
I	42%								58%	100%

Table 5
Source Categories Transition Probability Matrix of a Course Thesis Task

→	A	B	C	D	E	F	G	H	I	Σ
A		13%			13%			13%	63%	100%
B	25%	50%	25%							100%
C	13%	13%	50%	13%		13%				100%
D										100%
E		20%	20%		20%				40%	100%
F		100%								100%
G	100%									100%
H					50%				50%	100%
I		8%	8%		17%	8%	8%	8%	42%	100%

4.2 User Experiment Results

4.2.1 Expert's Search Trails Generation

According to the search-task relevant search trails generation method proposed in Section 3.1, the expert search trail case base for this experiment was generated and represented as follows:

$$\mathrm{TMC}=\{\mathbf{tmc}_1,\mathbf{tmc}_2\}$$

$$\mathbf{tmc}_1=\left(\mathbf{T}_{\mathrm{tmc}_1},\mathrm{Cat}_{\mathrm{tmc}_1},\mathrm{S}_{\mathrm{tmc}_1},\mathrm{SC}_{\mathrm{tmc}_1},\mathbf{A}_{\mathrm{tmc}_1}\right)$$

$$\mathbf{tmc}_2=\left(\mathbf{T}_{\mathrm{tmc}_2},\mathrm{Cat}_{\mathrm{tmc}_2},\mathrm{S}_{\mathrm{tmc}_2},\mathrm{SC}_{\mathrm{tmc}_2},\mathbf{A}_{\mathrm{tmc}_2}\right)$$

$\mathbf{T}_{\mathrm{tmc}_1}$ = (Class presentation, Low task complexity, Knowledge about "cloud computing")

$\mathbf{T}_{\mathrm{tmc}_2}$ = (Course thesis,High task complexity,Artificial intelligence in education or healthcare)

$\mathrm{Cat}_{\mathrm{tmc}_1}=\mathrm{Cat}_{\mathrm{tmc}_2}=\{A,B,C,D,E,F,G,H,I\}$, the categories of Internet information sources classified in Section 4.1.4.

$\mathrm{S}_{\mathrm{tmc}_1}$ and $\mathrm{SC}_{\mathrm{tmc}_1}$, the full search trails of the low complexity search task, were generated after expert searchers executed the task. For example, the Internet information sources accessed by *expert [5]* were baike.baidu.com, zhihu.com, xueshu.baidu.com, wanfangdata.com.cn, baidu.com, wiki.tw.lvfukeji.com, baidu.com, zhihu.com, cloud.idcquan.com, cloud.it168.com. Then, the Internet sources sequence $\mathbf{SrcCatSeq}_{\mathrm{expert[5]}}$ was stored in the $\mathrm{S}_{\mathrm{tmc}_1}$ set, and the corresponding $\mathbf{SrcCatSeq}_{\mathrm{expert[5]}}$ = (F,G,B,C,A,F,A,F,I,I) was stored in the $\mathrm{SC}_{\mathrm{tmc}_1}$ set. The transition probability matrix of the search trail Markov

Table 6
Search Trail Recommendation Evaluation

Task type	Trail type	Evaluation	Min	Max	Average	SD
Class presentation (Low task complexity)	Full trail	PU	2.7	6.3	4.7	1.30
		SA	3.0	7.0	5.1	1.18
		AC	4.0	7.0	4.9	1.03
	One-step trail	PU	3.0	6.3	4.6	1.08
		SA	3.0	7.0	5.0	1.27
		AC	3.7	7.0	5.0	1.12
Course thesis (High task complexity)	Full trail	PU	3.0	6.0	5.1	1.05
		SA	2.0	6.0	5.0	1.45
		AC	4.0	6.0	4.7	0.65
	One-step trail	PU	3.0	6.3	4.8	1.24
		SA	3.0	7.0	5.0	1.34
		AC	3.3	7.0	4.9	1.17

Table 7
Search Trail Recommendation Evaluation Clustering

Cluster no.	Class presentation (Low task complexity)						Course thesis (High task complexity)						Rating	N
	Full trail			One-step trail			Full trail			One-step trail				
	PU	SA	AC	PU	SA	AC	PU	SA	AC	PU	SA	AC		
1	5.8	6.0	6.2	5.8	6.3	6.3	5.8	5.8	5.5	6.0	6.3	6.2	High	2
2	4.7	5.2	4.5	4.4	4.8	4.6	5.3	5.4	4.5	4.6	4.8	4.5	Medium	4
3	2.7	3.0	4.0	3.0	3.0	4.0	3.0	2.0	4.0	3.0	3.0	4.0	Low	1

model of the class presentation task was estimated, as shown in Table 4.

Similarly, S_{tmc_2} and SC_{tmc_2}, the full search trails of the high complexity search task, were generated after expert searchers performed the task. For example, the Internet information sources accessed by *expert [1]* were baidu.com, tech.163.com, 360.doc.com, gov.cn, con.com.cn, ex.cssn.cn, useit.com.cn, xueshu.baidu.com, cnki.net, xueshu.baidu.com, baidu.com, blog.csdn.net, useit.com.cn, app.webofknowledge.com. Then, the Internet sources sequence $\mathbf{SrcCatSeq}_{expert[1]}$ was stored in the S_{tmc_2} set, and the corresponding $\mathbf{SrcCatSeq}_{expert[1]}$ = (A,I,E,I,I,I,E,B,C,B,A,H,E,C) was stored in the SC_{tmc_2} set. The transition probability matrix of the search trail Markov model of a course thesis task was estimated, as shown in Table 5.

Since the two types of search tasks in this study are highly differentiated, it is easy to get the expert search trails recommendation set of these search tasks as follows:

ET(Class presentation, Low task complexity,
Knowledge about "cloud computing") = $\{\mathbf{tmc}_1\}$

ET(Course thesis, High task complexity,
Artificial intelligence in education) = $\{\mathbf{tmc}_2\}$

ET(Course thesis, High task complexity,
Artificial intelligence in healthcare) = $\{\mathbf{tmc}_2\}$

4.2.2 Novice Searchers' Evaluation of the Search Trail Recommendation

From Table 6, it is evident that the average scores of novice searchers' perceptions (i.e., perceived usefulness (PU), satisfaction (SA), and acceptance (AC)) on the search trail recommendations are all higher than 4. This result shows that novice searchers could benefit from search trail recommendations and are willing to accept this search recommendation during the search process.

The extremely large score values for each measurement item indicate that different novice searchers may have different perceptions of the search trail recommendation. Therefore, we used the k-means clustering algorithm to cluster the scores (i.e., perceived usefulness, satisfaction, and acceptance). Here, the number of categories was set to 3. The clustering results are shown in Table 7.

According to the ratings, clusters 1, 2, and 3 are classified as high, medium, and low rating groups. Among the novice searchers, there were two participants in the high rating group, four participants in the medium rating group, and one participant in the low rating group. Then, we performed an in-depth analysis of the different rating groups. In this analysis, we considered the participants' background information, search behavior data, post-experiment questionnaire results, and data generated from interviews.

4.2.2.1 High-Rating Group

Novice [3] and *novice [7]* assigned a score greater than 5 to each evaluation item, indicating that they may have benefited significantly from the search trail recommendation during the search process. These novice participants said that the search trail recommendation was useful, enriched the diversity of information sources, and potentially, improved their search skills. They indicated that, before this experiment, they only used Baidu, the largest Chinese search engine, to search for information. They did not know that they could search for information in databases or professional forums.

The background survey showed that *novice [3]* only searched for academic information online once a week. In this experiment, *novice [3]* demonstrated that, during the search process, she had difficulty selecting information sources and forming queries.

Novice [7] stated that she only searched for academic information once a month, and she reported that her search skills were poor. She has difficulty expressing her information needs clearly. *Novice [7]* explained that "After reading the task requirements, I don't know how to make a search strategy and don't know which information source to search from at first. Fortunately, with the help of the search trail recommendation, I learned about many types of online information sources. In this experiment, the diversity of information sources and content in the search results is good, with greatly improved compared to before. I felt that my search skills had also improved."

4.2.2.2 Medium-Rating Group

The perception of usefulness, satisfaction, and acceptance scores for the search trail recommendation given by *novice [2]*, *novice [4]*, *novice [5]*, and *novice [6]* were all higher than 4. These novices indicated that the search trail recommendation helped them to increase the diversity of their searches.

The background survey showed that these students only searched for academic information online once every 3 days or once a month. *Novice [5]* felt that her information search skills were poor and that she had difficulty expressing her information needs clearly. However, she did not want to use a complex search strategy or change her search habits. *Novice [4]* and *novice [6]* also indicated low confidence in their information search abilities. However, *novice [4]* expressed willingness to use a complex search approach, and *novice [6]* expressed willingness to change their search strategy to suit different search tasks. Compared with other users, search results of *novice [5]* included relatively few information sources and low diversity of content. Participants in this cluster indicated that they usually obtained information from the Baidu search engine, Online encyclopedia, and E-journal databases. They also acknowledged that they had difficulty choosing information sources during the search process. For example, *novice [2]* said, "I don't know whether to use a search engine or go to a database;" *novice [4]* said, "I feel that the content I find on different pages is the same;" and *novice [6]* said, "I don't know which information source can find more information." These problems infer that these participants have no idea how to get diverse information.

4.2.2.3 Low-Rating Group

Novice [1] assigned a low score to perceptions of usefulness and satisfaction for both the full trail and the one-step trail recommendations for search tasks with high or low complexity. He assigned a score of 4 to the perception of acceptance, which indicated that he did not reject such recommendations. In the exit interview, *novice [1]* stated that the one-step trail recommendation could prompt

Table 8
Information Search Experience and Habits of Different Rating Clusters

Rating	No.	Information search experience and habits			
		Search frequency	Complex search intention	Search confidence	Change search strategy intention
High	3	**Medium**	**Medium**	**Medium**	Low
	7	Low	**High**	Low	Low
Medium	2	**Medium**	**High**	**High**	Low
	4	Low	**High**	**Medium**	Low
	5	Low	Low	Low	Low
	6	**Medium**	**Medium**	Low	**Medium**
Low	1	**High**	**Medium**	**High**	Low

him to obtain information from other information sources after obtaining information from some Internet sources. However, the recommendation of five full trails with different sequence patterns confused him and he did not know which one to choose.

The background survey showed that *novice [1]* searched for academic information online every day. He believes that he can clearly express his information needs and find the information easily. *Novice [1]* used various Internet sources in both search tasks, and his search results showed diversity. The difficulties he encountered in the search process were primarily related to queries, for example, "When searching for information about unfamiliar professional words, I could not find relevant pages in SERP by directly pasting its abbreviation in the search engine."

5 Discussion and Conclusion

In this paper, we have proposed a method to learn the behavior of expert searchers to support novice searchers engaged in the same or similar complex tasks. Developing the proposed method involves identifying the sequence of internet information sources explored by expert searchers using Markov chains and organizing these search experiences using a search-task relevant case base. We employ the CBR approach to generate search-task relevant search trail recommendations for novice searchers to assist them in identifying necessary steps or information sources to achieve task completion. We demonstrate through a user study that our task-relevant search trail recommendations can help improve novice searchers' search performance. The novice searchers who participated in the user study were satisfied with the search trail recommendations and were willing to use them in the future.

5.1 Discussion

The user study results showed that novice searchers have positive attitudes toward the search trail recommendation. Most novice searchers stated that the recommendation helped improve the diversity of information sources and the integrity of the search results' information content. They also indicated that they were satisfied with the search trail recommendation provided during the search process. All novice searchers were willing to accept the search trail recommendation.

An in-depth analysis of novice searchers' evaluations of the search trail recommendation found that searchers with different evaluation levels have different information search experience and habits (Table 8), encountered different problems in the search process, and have different task performance results (Table 9).

Participants with a high rating for search trail recommendations stated that the recommendation could help them to develop search strategies and select the category and quantity of information sources. Typically, they search online for academic information infrequently and are not confident in their search skills. During the information search process, they encountered difficulty understanding task requirements, developing an information search strategy, selecting information sources, and formulating queries. Participants in the medium rating group stated that the search trail recommendations could complement their source categories and increase the number of relevant sources during the search process.

Table 9
Task Performance of Different Rating Clusters

Rating	No.	Difficulties	Benefits	Search Results
High	3	Task understand, search strategies formulation, information sources section, queries	Develop a search strategy, select the type and number of sources	Good
	7	Task understand, search strategies formulation, queries		Good
Medium	2	Information sources section	Supplement source category and number	Good
	4	Search results evaluation		Good
	5	Task understand, search results evaluation		**Poor**
	6	Information sources section		Good
Low	1	Queries	Supplement sources	Good

They searched for online academic information every few days or once a month, with low search confidence. For this group, the difficulties encountered in the search process were primarily associated with information source selection and evaluation of the retrieved result. One novice searcher gave the search trail recommendation a low evaluation, which indicates that this recommendation did not significantly improve his information source diversity. This novice searcher searched for online information frequently, had a high degree of search confidence, and was willing to try a complex search. The only difficulties he encountered during the search process related to queries, and his search results were sufficient.

This experiment found that, related to the search trail recommendation, novice searchers who were willing to conduct complex search activities or to change search habits obtained information search results that demonstrated good information source diversity and content integrity. However, even though *novice [5]* was provided with search trail recommendations, her search results were relatively poor. This may have occurred because she has poor search skills and is unwilling to try complex searches or change her search habits.

Some novice searchers suggested that it would be more beneficial to include expert searchers' evaluation of each Internet source category in the search trail. This would allow them to know what information sources the expert searchers used and why the experts used these sources. Some novice searchers expect search engines to automatically separate aspects of an information search task. Besides, search engine result pages could be layered to display corresponding information source results, which would greatly reduce the searcher's workload and improve the effectiveness of search results.

5.2 Implications

The research presented in this paper could extend previous studies in several ways. First, the search-task relevant search trail recommendations proposed in this study expanded the scope of traditional trail recommendations, which primarily focus on query-relevant trail origins, sub-trails, and destination recommendations (White & Huang, 2010). Our findings suggest that task-level search trail recommendations will provide useful guidance to novice searchers. Second, we proposed a method for search trail generation using Markov chains and evaluated the method experimentally. The results suggested that Markov chains were useful tools for Internet information source sequence modeling and search trail analysis. Third, the most innovative feature of this study is that it recommended search trails based on CBR searching. Employing CBR searching represents a new research idea that may benefit research on Internet information recommendations for complex or exploratory search tasks.

The results of this study are also important for search engines and search assistance service designers. Our findings suggested that search trail recommendations incorporated into SERPs and certain webpage can help novice searchers. Most novice searchers reported good experiences interacting with the trails provided, and some even indicated that the search trail recommendations potentially improved their search skills. The information search trail recommendation method proposed in this study can provide a reference to improve network information organization and management, information recommendation services, and improve the utilization efficiency of professional or domain Internet sources.

5.3 Limitations and Future Work

Despite the theoretical and practical value of this research, we should acknowledge some limitations. First, the Markov chain model currently used to identify search trails is limited in terms of the amount of training data required and dimensionality with Internet sources categories classified in this study. In the future, more work needs to be done to extend the method to all Internet information sources.

Second, the use of a CBR approach in the search trail recommendation may also have adverse effects. The knowledge structures activated in the searcher's mind by the provided case may hinder their access to other areas of the solution space (Althuizen & Wierenga, 2014). In addition, this study covered a limited number of task features. Considering the various features that might exist in the actual CBR process, we will enrich features more comprehensively in the next study.

We have shown the important promise of our approach in supporting some important dimensions of search performance (i.e., develop a search strategy, select the type, and number of sources). However, we also need to evaluate the effectiveness of these recommendations with different evaluation criteria, such as relevance, topic coverage, and topic diversity of the search trail recommendations.

What we can infer from the search interaction log data is limited; however, our approach has provided insight on how to organize expert searchers' trails experience and how to generate and represent search-task relevant search trail recommendations on the SERPs and certain webpages. To provide more accurate and efficient information search trail recommendations, in the future, we intend to conduct large-scale user studies that consider and evaluate task and searcher characteristics, as well as Internet sources and queries.

Acknowledgments: This work is supported by the National Natural Science Foundation of China under Grant No. 71874130 and the Ministry of Education of the People's Republic of China under Grant No. 18YJC870026.

References

Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, *7*(1), 39–59. doi:10.3233/AIC-1994-7104

Alptekin, G. I., & Büyüközkan, G. (2011). An integrated case-based reasoning and MCDM system for Web based tourism destination planning. *Expert Systems with Applications*, *38*(3), 2125–2132. doi: 10.1016/j.eswa.2010.07.153

Althuizen, N., & Wierenga, B. (2014). Supporting creative problem solving with a case-based reasoning system. *Journal of Management Information Systems, 31*(1), 309–340. doi: 10.2753/MIS0742-1222310112

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

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *Management Information Systems Quarterly*, *25*(3), 351–370. doi: 10.2307/3250921

Bichindaritz, I., & Marling, C. (2006). Case-based reasoning in the health sciences: What's next? *Artificial Intelligence in Medicine*, *36*(2), 127–135. doi: 10.1016/j.artmed.2005.10.008

Bilenko, M., & White, R.W. (2008). Mining the search trails of surfing crowds: Identifying relevant websites from user activity. *Proceedings of the 17th International Conference on World Wide Web (WWW'08)*, 51–60. doi: 10.1145/1367497.1367505

Bridge, D., Göker, M. H., McGinty, L., & Smyth, B. (2006). Case-based recommender systems. *The Knowledge Engineering Review*, *20*(3), 315–320. doi: 10.1017/S0269888906000567

Capra, R., & Arguello, J. (2019). Using trails to support users with tasks of varying scope. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'19)*, 977–980. /doi: 10.1145/3331184.3331290

Gavetti, G., & Rivkin, J. W. (2005). How strategists really think: Tapping the power of analogy. *Harvard Business Review*, *83*(4), 54–63. doi: 10.1177/105901103257989

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*(2), 155–170. doi: 10.1207/s15516709cog0702_3

Goldstein, D. K. (2001). Product managers' use of scanner data: A story of organizational learning. In R. Deshpande (Ed.), *Using Market Knowledge* (pp. 164–191). Los Angeles, USA: SAGE Publishing.

Gregan-Paxton, J., & Cote, J. (2000). How do investors make predictions? Insights from analogical reasoning research. *Journal of Behavioral Decision Making*, *13*(3), 307–327. doi: 10.1002/1099-0771(200007/09)13:33.0.CO;2-3

Hassan, A., & White, R. W. (2012). Task tours: Helping users tackle complex search tasks. *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM'12)*, 1885–1889. doi: 10.1145/2396761.2398537

He, W., Erdelez, S., Wang, F.-K., & Shyu, C.-R. (2008). The effects of conceptual description and search practice on users' mental models and information seeking in a case library with a best match search mechanism. *Information Processing & Management*, *44*(1), 294–309. doi: 10.1016/j.ipm.2007.03.008

He, W., & Tian, X. (2017). A longitudinal study of user queries and browsing requests in a case-based reasoning retrieval system. *Journal of the Association for Information Science and Technology*, *68*(5), 1124–1136. doi: 10.1002/asi.23738

Hendahewa, C., & Shah, C. (2017). Evaluating user search trails in exploratory search tasks. *Information Processing & Management*, *53*(4), 905–922. doi: 10.1016/j.ipm.2017.04.001

Huang, J., Wang, H., Zhang, W., & Liu, T. (2020). Multi-task learning for entity recommendation and document ranking in web search. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *11*(5), 1–24. doi: 10.1145/3396501

Hüllermeier, E. (2007). *Case-based approximate reasoning*. Netherlands: Springer. doi: 10.1007/1-4020-5695-8

Joachims, T., Freitag, D., & Mitchell, T. (1997). WebWatcher: A tour guide for the World Wide Web. *Proceedings of the 15th International Joint Conference on Artificial Intelligence*, 770–775. doi: 10.1063/1.2339014

Kim, J. (2009). Describing and predicting information-seeking behavior on the Web. *Journal of the American Society for Information Science and Technology*, *60*(4), 679–693. doi: 10.1002/asi.21035

Li, Y. (2009). Exploring the relationships between work task and search task in information search. *Journal of the Association for Information Science and Technology*, *60*(2), 275–291. doi: 10.1002/asi.20977

Li, Y., & Belkin, N. J. (2010). An exploration of the relationships between work task and interactive information search behavior. *Journal of the American Society for Information Science and Technology*, *61*(9), 1771–1789. doi: 10.1002/asi.21359

Lin, W. S., & Wang, C. H. (2012). Antecedences to continued intentions of adopting e-learning system in blended learning instruction: A contingency framework based on models of information system success and task-technology fit. *Computers & Education*, *58*(1), 88–99. doi: 10.1016/j.compedu.2011.07.008

Liu, Y., Huang, X., & An, A. (2007). Personalized recommendation with adaptive mixture of Markov models. *Journal of the American Society for Information Science and Technology*, *58*(12), 1851–1870. doi: 10.1002/asi.20631

Moon, H. S., Kim, J. K., & Ryu, Y. U. (2013). A sequence-based filtering method for exhibition booth visit recommendations. *International Journal of Information Management*, *33*(4), 620–626. doi: 10.1016/j.ijinfomgt.2013.03.004

Moore, J., Erdelez, S., & He, W. (2006). Retrieval from a case-based reasoning database. *Academic Exchange Quarterly*, *10*(4), 65–68.

O'Day, V. L., & Jeffries, R. (1993). Orienteering in an information landscape: How information seekers get from here to there. *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI'93)*, 438–445. doi: 10.1145/169059.169365

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

Pandit, S., & Olston, C. (2007). Navigation-aided retrieval. *Proceedings of the 16th International Conference on World Wide Web (WWW'07)*, 391–400. doi: 10.1145/1242572.1242626

Piccialli, F., Cuomo, S., Giampaolo, F., Casolla, G., & di Cola, V. S. (2020). Path prediction in IoT systems through Markov Chain algorithm. *Future Generation Computer Systems, 109*, 210–217. doi: 10.1016/j.future.2020.03.053

Pirolli, P., & Card, S. K. (1999). Information foraging. *Psychological Review*, *106*(4), 643–675. doi: 10.1037/0033-295X.106.4.643

Sarukkai, R. R. (2000). Link prediction and path analysis using Markov chains. *Computer Networks*, *33*(1-6), 377–386. doi: 10.1016/S1389-1286(00)00044-X

Shokouhi, S. V., Skalle, P., & Aamodt, A. (2014). An overview of case-based reasoning applications in drilling engineering. *Artificial Intelligence Review*, *41*(3), 317–329. doi: 10.1007/s10462-011-9310-2

Singla, A., White, R., & Huang, J. (2010). Studying trailfinding algorithms for enhanced web search. *Proceedings of the 33rd international ACM SIGIR conference on Research and Development in Information Retrieval (SIGIR'10)*, 443–450. doi: 10.1145/1835449.1835524

Smith, C. L., Gwizdka J., & Field, H. (2016). Exploring the use of query auto completion: Search behavior and query entry profiles. *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval (CHIIR'16)*, 101–110. doi: 10.1145/2854946.2854975

Tabatabai, D., & Shore, B. M. (2005). How experts and novices search the Web. *Library & Information Science Research*, *27*(2), 222–248. doi: 10.1016/j.lisr.2005.01.005

Trigg, R. H. (1988). Guided tours and tabletops: Tools for communicating in a hypertext environment. *ACM Transactions on Information Systems*, *6*(4), 398–414. doi: 10.1145/58566.59299

Wexelblat, A., & Maes, P. (1999). Footprints: History-rich tools for information foraging. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'99)*, 270–277. doi: 10.1145/302979.303060

Wheeldon, R., & Levene, M. (2003). The best trail algorithm for assisted navigation of web sites. *Proceedings of the First Conference on Latin American Web Congress (LA-WEB'03)*, 166–178. doi: 10.1109/LAWEB.2003.1250294

White, R.W., Bilenko, M., & Cucerzan, S. (2007). Studying the use of popular destinations to enhance web search interaction. *Proceedings of the 30th annual international ACM SIGIR conference on Research and Development in Information Retrieval (SIGIR'07)*, 159–166. doi:10.1145/1277741.1277771

White, R. W., Dumais, S. T., & Teevan, J. (2009). Characterizing the influence of domain expertise on web search behavior. *Proceedings of the Second International Conference on Web Search and Web Data Mining (WSDM'09)*, 132–141. doi: 10.1145/1498759.1498819

White, R. W., & Huang, J. (2010). Assessing the scenic route: measuring the value of search trails in web logs. *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'10)*, 587–594. doi: 10.1145/1835449.1835548

Yuan, X., & White, R. (2012). Building the trail best traveled: Effects of domain knowledge on Web search trailblazing. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'12)*, 1795–1804. doi: 10.1145/2207676.2208312

Ziegler, C.-N., McNee, S. M., Konstan, J. A., & Lausen, G. (2005). Improving recommendation lists through topic diversification. *Proceedings of the 14th International Conference on World Wide Web (WWW'05)*, 22–32. doi: 10.1145/1060745.1060754