Physica A 506 (2018) 844–856

Contents lists available at ScienceDirect

Physica A

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

Towards a coherence-oriented complex search experience management method

Yin Zhang [a], Bin Zhang [a,*], Kening Gao [a], Pengfei Li [a], Yuli Zhao [b], Changsheng Zhang [a]

[a] School of Computer Science and Engineering, Northeastern University, Shenyang, China
[b] Software College, Northeastern University, Shenyang, China

HIGHLIGHTS

- Search management systems should maintain temporal, causal and thematic coherence.
- We propose a new method to maintain the 3 types of coherence for complex search.
- We propose the relative chronological source-tracking tree to represent search task.
- We conducted a user study to evaluate the method in 2 types of complex search task.

ARTICLE INFO

Article history:
Received 2 November 2017
Received in revised form 11 February 2018
Available online 7 May 2018

Keywords:
Complex search
Search experience
Consistence-oriented
TimeTree

ABSTRACT

Experiences of complex search tasks are important in social interaction and in problem solving. Considering the high importance of complex search experiences, many search experience management systems (SEMSs) have been introduced. Like any other life experience, complex search experiences should maintain 3 types of global coherence: temporal, causal and thematic coherence. However, to the best of our knowledge, none of the available SEMSs were designed to support all the 3 types of global coherence. In this paper, we introduce a coherence-oriented complex search experience management method named TimeTree. By organizing queries and clicks of a complex search task as a relative chronological source-tracking tree (RCST), TimeTree manages to support all the 3 types of global coherence. We describe a user study to evaluate TimeTree in 2 typical types of complex search task. The subjective evaluation results, the expert evaluation results, and the objective evaluation results all suggest that TimeTree can help maintain temporal, causal and thematic coherence for complex search experiences.

1. Introduction

Web users are using search engines to solve tasks with increasing complexity. Examples of complex search tasks could be collecting related papers for a survey, or planning a holiday trip. The cost of a complex search task could be very high, both in cognition and in time [1]. In terms of cognition cost, a complex search task may contain many queries and clicks. Thus, a complex search task may lead to the compression of many concepts and documents. Then considering the time cost, complex search tasks could also be very time consuming. A user may need several days, weeks or months to complete a

* Corresponding author.
E-mail addresses: zhangyin@cse.neu.edu.cn (Y. Zhang), zhangbin@cse.neu.edu.cn (B. Zhang), gaokening@cse.neu.edu.cn (K. Gao).

https://doi.org/10.1016/j.physa.2018.03.036

complex search task. The high cost of cognition and time makes complex search tasks hard to complete. Meanwhile, the high cost also makes the experience of how to complete a complex search task valuable.

Experiences of complex search tasks are important in social interaction and in problem solving [2]. For example, someone may need to explain how he searched to solve a complex problem to his colleagues. Or, someone may need to apply the experiences of some past search tasks to new problem settings. Complex search experiences are also important in research. Many methods have also been proposed to mine knowledge from previous search experiences to support current search tasks [3,4]. Considering the high importance of complex search experiences, many search experience management systems (SEMSs) were proposed to help searchers organize and manage search experiences [1,5–9].

Like any other life experience, complex search experiences should maintain 4 basic types of global coherence: temporal, causal, thematic, and cultural concept of biography [10]. Since search tasks will not last long enough for the cultural concept of biography to change, we argue that complex search experiences should maintain at least 3 basic types of global coherence: temporal, causal and thematic. For example, consider a student working on a report about chemotherapy drugs. The student is asked to explain how he searched to find all the information. For a narrative of the search task, he first needs to create temporal coherence. He needs to reconstruct the temporal order of the queries and the clicks performed, such as ''I first Binged 'drugs in chemotherapy', and I found a list of cancer chemotherapy drugs from navigatingcare.com. Then I found a page talking about how chemotherapy drugs work from cancer.org''. The student also needs to explain why he searched for a new query, such as ''I learnt that keyword from the cancer.org page'' or ''The new query is a reformulation of the previous one''. The explanation is necessary to make his actions reasonable, which achieves causal coherence. Finally, the student needs to explain how he divided the complex search task into subtasks, such as ''These are queries and clicks about alkylating agents and those are about antimetabolites''. A user usually divides a complex information need into several sub-tasks of small topics [11]. These sub-tasks establish thematic coherence to provide a clear view of the complex search task.

However, to the best of our knowledge, none of the available SEMSs were designed to support all the 3 types of global coherence. Such an inability makes these SEMSs unable to properly capture complex search experiences, and may cause negative impacts on the application of complex search experiences in social interaction, problem solving and in research. Thus, new designs should be proposed to overcome such an inability.

In this paper, we introduce a coherence-oriented complex search experience management method named TimeTree. TimeTree organizes queries and clicks of a complex search task as a relative chronological source-tracking tree (RCST). An RCST enables a user to quickly identify facts about the 3 types of global coherence for queries and clicks. Moreover, TimeTree also facilities supplementary tools to enable rating and commenting on queries and clicks. The RCST structure and the supplementary tools give TimeTree the ability to properly maintain temporal, causal and thematic coherence for complex search experiences.

We organize this paper as follows: We review the available SEMSs and other research works related to complex search experiences in Section 2. In Section 3 we introduce our complex search experience management method TimeTree, as well as how TimeTree maintains temporal, causal and thematic coherence for complex search experiences. To verify whether TimeTree could maintain all the 3 types of global coherence, we designed a user study in Section 4. We then conduct subjective evaluations, expert evaluations and objective evaluations for the user study in Section 5.

2. Related work

We discuss 3 areas of related work that are relevant to the research of this paper: (1) supporting complex search, (2) managing complex search experiences, and (3) managing other types of digital experiences.

Supporting complex search: Supporting complex search is an old problem dated back to the 1970s [12]. Recently, increasing attention has been paid to supporting complex search. Good surveys and reports on supporting complex search could be found at [13–15] and [16]. The problem of supporting complex search is a complex problem. We could identify lots of research topics in supporting complex search, such as task complexity [17], intent modeling [8,18], subtask recommendation [19,20], and system designs [21,22]. However, all these research topics seem to be supporting different aspects of one key problem. That key problem is to decompose a complex information need into smaller parts [14]. For this paper, the key to managing complex search experiences is also to manage experiences of how to decompose complex information needs. We provide temporal, causal cues and thematic supports to help a user manage experiences of how to decompose complex information needs.

Managing complex search experiences: In this paper, we introduce our complex search experience management method TimeTree. Before TimeTree, several SEMSs have been proposed. The SearchBar system [1] is the most like TimeTree. SearchBar works as an Internet Explorer plugin. SearchBar organizes queries and the corresponding clicks under user-entered topics. By these means, SearchBar enables task context resumption and information re-finding. However, the ability of SearchBar to maintain the 3 types of global coherence for complex search experiences is very lack. SearchBar does not have an intuitive way to show the chronological order of queries and clicks of a complex search task. It also lacks the ability to track the provenances of queries. Moreover, the plain topic structure also makes SearchBar unable to reflect the complex topical structures of complex search tasks. These disadvantages limit the performance of SearchBar in managing complex search experiences.

Besides SearchBar, several other search task management systems are also related to TimeTree. The sketchBrain system [7] keeps tracks of queries and post-query click streams as graphs. SketchBrain also supports operations on queries

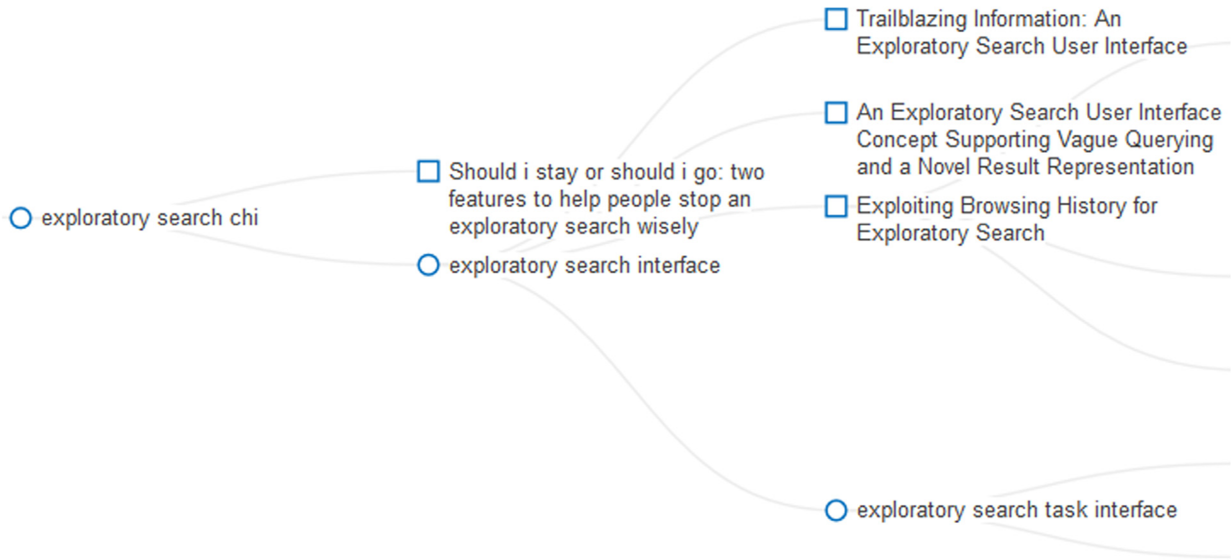


Fig. 1. A fraction of a TimeTree for a complex search task. Queries are represented as circle nodes. Clicks are represented as square nodes.

and click streams, such as annotation, projection, selection and classification. But, like SearchBar, sketchBrain also lacks intuitive ways to present the chronological order of queries and clicks as well as to track the provenances of queries. The IntentStreams system [5] supports parallel browsing and branching during search. IntentStreams provides a horizontally scrollable workspace, with queries and search results listed vertically. Again, a user could not easily tell the provenances of queries and the topic structures of complex search tasks. The SensePath system [23] provides a timeline based view to display captured sensemaking actions in their temporal order. The design of SensePath is to record the sensemaking processes of web searchers to facilitate HCI research. The lack of structural information of the recorded sensemaking actions makes SensePath hard to help recall subtask structure when recalling search tasks. [6,24–26] and [27–29] also suggested different complex search experience management methods. However, these systems all have limitations in supporting some or all of the 3 types of global coherence for complex search experiences.

Supporting other types of digital experiences: Experiences of past events are important in social interaction and in problem solving. [30] showed that personal experiences can help a speaker to educate and inform listeners. [2] suggested that the general or schematic knowledge abstracted from past experiences may not be relevant to a current problem. Thus, to solve the current problem, searching back through autobiographical memory to find a specific experience where a similar problem was encountered may be more useful. Thus, we believe that experiences of complex search tasks are also important in social interaction like imparting search experiences, and in problem solving using search engines.

Various techniques have been proposed to manage different types of digital experiences. van den Hoven et al. [31] designed a device to support recollecting personal experiences. They found that certain types of experience triggers may reduce the number of experiences. Murakami et al. [32] presented a method to create a user knowledge space from web searches, twitter, emails, calendars, and book purchases. They conducted a user study to evaluate the usefulness of the system. Sun et al. [33] presented a model of the posting behaviors in news forums considering the collective memory. The interest in life logging devices also leads to new methods in managing digital experiences [34,35].

3. Managing complex search experiences with TimeTree

In this section, we first introduce our proposed complex search experience management method TimeTree. We then give the implementation details of TimeTree. Finally, we introduce how TimeTree maintains temporal, causal and thematic coherence for complex search experiences.

3.1. The complex search experience management method TimeTree

To help searchers management complex search experiences while maintaining temporal, causal and thematic coherence, we introduce TimeTree. Each search task is represented as a TimeTree. A fraction of a TimeTree is shown in Fig. 1.

In the TimeTree of a complex search task, a query is represented as a circle node labeled with the corresponding query keywords. For example, the circle node to the middle-left of Fig. 1 labeled with "exploratory search chi" is a query node. A click is represented as a square node labeled with the corresponding result title. For example, the square node in the middle

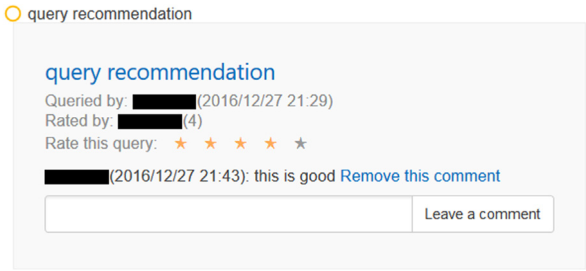


Fig. 2. An example of a query node infobox.

of Fig. 1 is a click node corresponding to a clicked search result entitled "Should I stay or should I go: two features to help people stop an exploratory search wisely".

We use arcs to connect nodes in a TimeTree to track the provenance of nodes. The direction of an arc is always from left to right. That means the node on the left side of an arc is always the source node, and the node on the right side is always the target node. Here we follow the definitions of source node, target node and sibling node of directed graphs. An arc exists between two nodes if:

(1) the source node is a query node, the target node is a click node, and the click is made in the search engine result page (SERP) of the query;
(2) the source node is a click node, the target node is a query node, and the provenance of the query is the clicked search result;
(3) the source node and the target node are both query nodes, and the target query node is a post query of the source query node.

In a TimeTree, a node can have multiple out-going arcs, but only one in-going arc. By these means, the nodes and the arcs of a TimeTree form a source-tracking tree structure. Nodes are then arranged in relative chronological order:

(1) The source node of an arc has an earlier timestamp than the target node, and is located on the left side of the arc. This means an ancestor node in a TimeTree always has an earlier timestamp than a descendant node.
(2) A sibling node with an earlier timestamp is always located above the other sibling with later timestamps. The parent node of some sibling nodes is located to the center-left of all the sibling nodes.

By arranging a source-tracking tree in a relative chronological order, we form a relative chronological source-tracking tree (RCST) structure for a TimeTree.

Detailed information about a node is provided in an infobox. Fig. 2 shows an example of an infobox. A user could see the infobox of a node by hovering the mouse pointer over the node. A user could rate, comment on, or delete the node in the infobox. A higher rating will brighten up the node.

3.2. Implementation details

Like SearchBar [1], TimeTree is implemented as an add-on for Mozilla Firefox. After installed, TimeTree is available as a browser window sidebar, as shown in Fig. 3. The TimeTree sidebar opens automatically when a user visits Google, Bing, or Baidu (the most frequently used search engine in China). The user could hide the TimeTree sidebar by clicking on the TimeTree toolbar button or the "Hide" button. The user could open the content of the TimeTree sidebar in a separate browser tab by clicking on the "Open in tab" button. The user could also open multiple browser windows. Each browser window has its own TimeTree sidebar. However, the contents of all the TimeTree sidebars are synchronized.

TimeTree monitors queries and clicks performed on the opened search engine. When a user submits a new query, a query node labeled with the query words is created in the sidebar. When a user clicks on a new search result, a click node labeled with the result title is created.

A previously submitted query or a previously clicked result will not lead to a new node. Instead, the content of the TimeTree sidebar is centered on the existing node corresponding to the query or click.

TimeTree creates an arc for a new query or click by tracking the source of the new query or click. For a new click, TimeTree identifies the search engine results page (SERP) where the click came from. TimeTree then considers the query of the SERP as the source of the new click. While for a new query, TimeTree assumes the user reads the query from the content of a click. Currently TimeTree assumes the user reads the query from the click with the highest TF/IDF score concerning the query. If such a click exists, the click is considered as the source of the new query. If such a click does not exist, the query submitted before the new query is considered as the source.

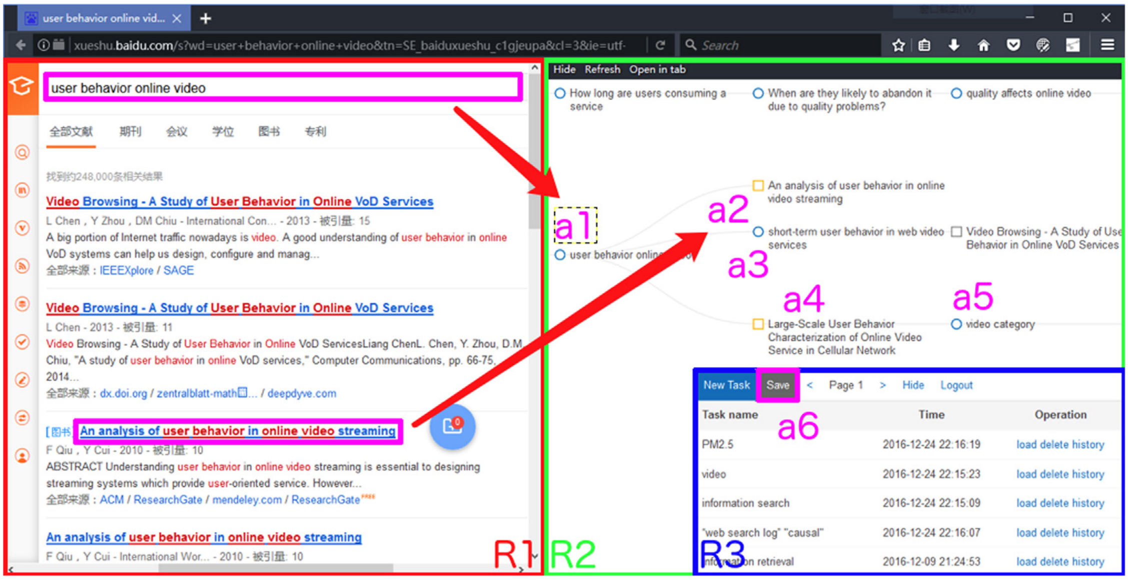


Fig. 3. A screenshot of TimeTree running in Mozilla Firefox. The left pane (R1) is a standard browser tab showing search results returned by a search engine. The right pane (R2) is the TimeTree sidebar, showing queries and clicks of the current search task. The queries and the clicks are organized in an RCST structure. The bottom-right pane (R3) in the TimeTree sidebar allows the current user to switch between different search tasks.

We use an example to show how a user uses TimeTree. The example consists of 6 actions, a1 to a6, as shown in Fig. 3. For action a1, the user enters a query "user behavior online video" in the browser tab (R1). After the user hits enter, a query node entitled "user behavior online video" is created in the TimeTree sidebar (R2). For action a2, the user clicks on a result in R1, and the corresponding click node is created in R2 entitled with the title of the result. Then for action a3, the user issues a new query "short-term user behavior in web video services". As TimeTree fails to identify the provenance click node for query a3, it is treated as a post query of query a1. Meanwhile, as the timestamp of query a3 is latter than click a2, query a3 is located below click a2. Then, the user may go back to the SERP of query a1 and makes a click action a4. Since click a4 happens in the SERP of query a1, it is treated as a sibling of a2 and a3 and is located below query a3 as it has the latest timestamp. Then the user issues a new query a5. At this time, TimeTree notices that click a4 may be the provenance of query a5 and locates a5 as a child of a4. The user then clicks on the "Save" button to save a snapshot of the current search task.

In the TimeTree sidebar, a new query or click node is added to the tree structure as a child of its source node. A user could adjust the tree structure by giving a node a new parent. To give a node a new parent, a user drags a node over the new parent node. TimeTree then disconnects the node from its old parent and moves it to its new location. A node is always located to the right of its parent. Sibling nodes are positioned vertically by time. Nodes at the same level are vertically aligned.

A task is named as the root query node (i.e., the first query node) of the task. A user could also create, save, load, delete or see 10 history versions of a search task in the TimeTree sidebar. A history version is a snapshot of a search task. Every time a user clicks on the "Save" button, a history version is created. A user could click on the "history" link in the bottom-right pane of the TimeTree sidebar to show and switch between all the history versions of a search task.

3.3. Maintaining global coherence for complex search experiences with TimeTree

Managing complex search experiences is closely related to the research problems in cognitive psychology [36]. Like any other kind of life experiences, complex search experiences should maintain 4 types of global coherence: temporal coherence, cultural concept of biography, causal coherence, and thematic coherence [10]. However, search tasks will not last long enough for the cultural concept of biography to change. Thus, we argue that complex search experiences should maintain at least 3 of the 4 coherences without the cultural concept of biography.

In this section, we discuss how TimeTree maintaining global coherence for complex search experiences. In detail, we discuss how a user achieves temporal coherence, causal coherence and thematic coherence for complex search experiences, as well as how TimeTree helps a user achieve these three types of coherence.

We first consider the causal coherence. A user achieves causal coherence by explaining the reason he submits a query or clicks a search result, such as "I learnt that keyword from the cancer.org page" or "The new query is a reformulation of the previous one". During a search process, a user could take notes on the reasons of his queries and clicks. However, note-taking induces more cognitive load [37]. [1] also showed that note-taking is a less popular feature in the search task management system SearchBar. Thus, in TimeTree, we introduce a vaguer but simpler way to help a user remember the reason for a query or a click. The way is to track the source of a query or a click, namely source-tracking, as introduced in Section 3.1. In contrast to note-taking, source-tracking requires no more than 1 drag-n-drop operation to complete. Source-tracking clearly induces less cognitive load. Meanwhile, although source-tracking only provides simple and vague cues on the reasons of queries and clicks, experimental results show that it is an effective way to maintain causal coherence in managing complex search experiences.

We then consider the thematic coherence. A user creates thematic coherence by interrelating queries and clicks by topics, such as "These are queries and clicks about alkylating agents and those are about antimetabolites". When performing a complex search task, a user usually divides the complex information need into several sub-tasks of small topics [11]. The user completes the complex search task by completing each sub-task individually. The queries and the clicks of a sub-task are interrelated by the topic of the sub-task. These sub-tasks establish thematic coherence to provide a clear view of the complex search task. In TimeTree, query nodes and click nodes are interrelated by source-tracking. Starting from a node n, source-tracking helps a user organize all the follow-up queries and clicks of the same topic of node n as a subtree rooted at n. Such subtree structures make it easy for the user to identify sub-tasks and maintain thematic coherence.

We finally discuss the temporal coherence. Temporal coherence refers to the temporal order of the queries and the clicks performed, such as "I first Binged 'drugs in chemotherapy', and I found a list of cancer chemotherapy drugs from navigatingcare.com. Then I found a page talking about how chemotherapy drugs work from cancer.org". However, such temporal order does not mean to tell the chronological order of any two queries or clicks. As mentioned before, a user usually divides a complex information need into several sub-tasks. In this case, the user only needs to maintain the temporal order of the queries and the clicks performed within the scope of each sub-task. Then the user only need to maintain the temporal order of each sub-task. TimeTree organizes query nodes and click nodes as an RCST. In an RCST, the descendant nodes of a node are always located to the right of the node. The follow-up sibling nodes of a node are always located below the node. With an RCST, a user could easily determine the relative chronological order of nodes according to their locations. Thus, the RCST of TimeTree could help a user maintain temporal coherence of complex search tasks.

4. User study design

In this section, we present the user study we designed to verify whether TimeTree could maintain all the 3 types of global coherence.

The user study was to understand whether and to what extent the RCST of TimeTree could help a user achieve temporal coherence, causal coherence and thematic coherence for complex search experiences. We achieved this goal by comparing TimeTree with a baseline SEMS we designed named SearchLog. SearchLog maintained limited global coherence for complex search experiences. SearchLog maintained temporal coherence, but has obvious disadvantages in maintaining causal and thematic coherence. As SearchLog did help maintain temporal coherence, TimeTree should outperform SearchLog at least in metrics about causal and thematic coherence. Meanwhile, SearchLog should not outperform TimeTree in metrics about temporal coherence.

We designed 4 complex search tasks as presented in Section 4.2. The tasks required skilled searchers. The tasks also required participants to be native Chinese speakers with similar English skills. Therefore, we recruited 32 postgraduate students majored in computer science from Northeastern University, China. Each participant is paid 60 CNYs (about 8.6 USDs) per hour.

We asked the participants to perform the 4 search tasks with TimeTree and SearchLog. We logged all the search operations performed by the participants. We also logged all the eye-gaze data. After 2 weeks, we asked the participants to share their search experiences by presenting how they conducted the search tasks. Search tasks performed with TimeTree and SearchLog are presented with TimeTree SearchLog accordingly. Details about the baseline system SearchLog is presented in Section 4.1. During presenting, we also logged all the eye-gaze data in TimeTree and SearchLog. Details about the study procedure is presented in Section 4.3.

4.1. The baseline SEMS SearchLog

We built a system named SearchLog as a baseline for comparison with TimeTree in managing complex search experiences. Like TimeTree, SearchLog also monitors queries and clicks performed on Google, Bing and Baidu. However, SearchLog only provides a hierarchical list of visited search results according to the search queries leading to the results. SearchLog presents the list in a separate Web page. Beside that page, SearchLog provides no other interface or functionality. A screenshot of SearchLog is given in Fig. 4. The SearchLog interface was used when the participants presenting their search tasks. The interface was not shown to the participants when searching.

SearchLog maintains limited global coherence for complex search experiences. SearchLog can maintain temporal coherence by organizing queries and clicks in a chronological ordered list. SearchLog could also maintain limited causal coherence by organizing clicks according to queries. However, SearchLog provides no information about the provenance of queries. Meanwhile, a user could hardly tell the sub-task structure of a complex task from SearchLog. Thus, SearchLog has obvious disadvantages in maintain causal and thematic coherence for complex search experiences.

By comparing TimeTree with SearchLog, we expected to see TimeTree outperformed SearchLog at least in metrics about causal and thematic coherence. Meanwhile, SearchLog should not outperformed TimeTree in metrics about temporal coherence.

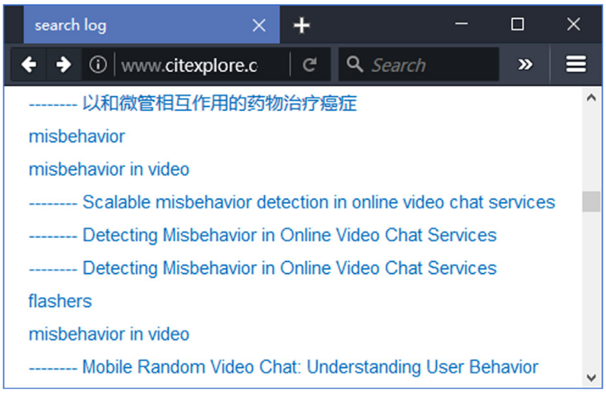


Fig. 4. A screenshot of SearchLog. Queries are presented with no intent. Clicks are indented with 8 hyphens.

4.2. Search task

In the user study, we wanted the participants to carry out complex search tasks. We adopted the definition of complex tasks proposed in [17] saying "A search task is complex if it requires at least one of the elements aggregation, discovery and synthesis". We also considered the three situations proposed in [38] in which users may conduct complex (exploratory) search tasks: (1) they lack the knowledge or contextual awareness to formulate queries or navigate complex information spaces, (2) the search task requires browsing and exploration, or (3) system indexing of available information is inadequate. We conducted the user study on scholar search engines. In scholar search engines, a full-text index may not be available for a paper in many cases. Thus, situation 3 lays in the nature of scholar search engines. Therefore, we only considered situation 1 and 2 when designing complex search tasks. We designed 4 search tasks. The first 2 tasks correspond to situation 2 and are called type 1 tasks. The later 2 tasks correspond to situation 1 and are called type 2 tasks.

For task type 1, we asked participants to write reports according to task descriptions. To write the reports, the participants had to collect several aspects of information according to the task descriptions. We confirmed that all the participants have no professional knowledge about the topics of the task descriptions. By these means, we set up 2 tasks that "requires browsing and exploration". The 2 tasks all consist of several subtasks with open answers. Since the volunteers have no professional knowledge about the tasks, then will have to start from some general searches, reading documents and digging into possible directions. Thus, these tasks satisfy situation 2. To maximize the effort of browsing and exploration, type 1 tasks were described and were completed in Chinese. The translated descriptions of the two type 1 tasks are given as below:

Task 1, Type 1: Write a report about drugs used in chemotherapy. Introduce common drugs used in chemotherapy, including their pharmacological actions, indications, administrations, and side effects. Introduce treatment strategies and common combination chemotherapy regiments. Do not focus only on drugs that beat cancer. Consider also drugs that decrease the toxic effects of other drugs.

Task 2, Type 1: Write a report about fine particles (PM 2.5) in China. Introduce the concept and sources of PM 2.5. Introduce how and why PM 2.5 affects human health. Introduce the top affected cities by PM 2.5 in China. Analyze the causes of PM 2.5 in the top affected cities. Introduce methods as well as the corresponding mechanisms and feasibilities to reduce PM 2.5. Analyze both the positive and the negative effects of the introduced methods.

For task type 2, we first gave participants some computer science research materials to read without telling them what were the tasks. After 2 days, we asked the participants to search for papers related to the materials merely based on their memories. We asked each participant to find a list of related papers for each material. We confirmed that all the participants have understood the materials after they read the materials so that they understand the information need. We also confirmed that the research topics of all the participants do not match the topic of the materials so that they "lack the knowledge or contextual awareness to formulate queries for navigate complex information spaces". Since the research topics of the volunteers did not match the topics of the materials, the knowledge is limited for the volunteers. When the experiment started, the volunteers were all surprised about the tasks and declared that they had forget the details or even the topics of the materials. The ambiguous memories limited the contextual awareness of the volunteers. In these situations, the volunteers would have to make many tentative searches and evaluate search results with ambiguous relative measures. Thus, these tasks satisfy situation 1. To increase difficulty in formulating queries, type 2 tasks were written and are completed in English. The two materials were the abstracts of the following papers:

Task 3, Type 2: S. Shunmuga Krishnan, Ramesh K. Sitaraman: Video Stream Quality Impacts Viewer Behavior: Inferring Causality Using Quasi-Experimental Designs. IEEE/ACM Trans. Netw. 21(6): 2001–2014 (2013).

Task 4, Type 2: Hanqiang Cheng, Yu-Li Liang, Xinyu Xing, Xue Liu, Richard Han, Qin Lv, Shivakant Mishra: Efficient misbehaving user detection in online video chat services. WSDM 2012: 23–32.

Table 1
The 6 parts of the user study for each participant.

Part #	Conducted on	Works to do
1	The 1st day	(1) Learn to use TimeTree for about 10 min. (2) Read the materials of the type 2 tasks. (3) Perform the type 1 tasks for about 25 min each.
2	The 3rd day	Perform the type 1 tasks for about 25 min each. (2) Perform the type 2 tasks for about 25 min each.
3	The 5th day	Perform the 4 tasks for about 25 min each.
4	The 8th day	(1) Write reports for type 1 tasks. (2) Make lists for type 2 tasks.
5	The 10th day	Take (1) a survey and (2) an interview about the search experience.
6	The 19th day	(1) Make presentations on how he conducted the 4 tasks. (2) Take a survey about the recollection experience.

4.3. Study procedure

The user study was to understand how and to what extent TimeTree and SearchLog help searchers maintain global coherence for complex search experiences. To do this, we asked the participants to conduct the 4 search tasks introduced in Section 4.2. After the participants finished the search tasks, they were asked to present their search process to share their search experiences. Thus, we could divide the study into two main blocks: the searching block, and the presenting block.

We first considered the design of the searching block. According to [1], a complex search task usually spans over several hours or days. During this time, the search task may be interrupted several times. To simulate this situation, we divided the searching block into 4 parts spanning over 8 days. Compared with a single session of complex search, we believed such interrupted search sessions better reflected the situations a searcher confronted in real-life complex search tasks.

We then considered the design of the presenting block. In the presenting block, the participants shared their search experiences by presenting their search tasks. If the participants shared their search experiences immediately after they finished the search tasks, they might present their search tasks based on their short-term memories instead of TimeTree and SearchLog. However, we wanted the participants to reconstruct their search experiences using TimeTree and SearchLog, instead of using their short-term memories. Thus, according to the forgetting curve of Ebbinghaus, the presentations were arranged 10 days after the searching block. By these means, we believed the participants would have to reconstruct search experiences based on TimeTree and SearchLog instead of using their short-term memories.

We asked each participant to perform all the 4 tasks proposed in Section 4.2. For each participant, we randomly chose a type 1 task for him to complete using TimeTree. Then, the other type 1 task is completed using SearchLog. The same strategy is also applied to the type 2 tasks. We managed to have each task performed by 16 participants on TimeTree, and by the other 16 participants on SearchLog.

For each participant, the user study consisted of 6 parts. The 6 parts were conducted on 6 different days. Details about the 6 parts are shown in Table 1. On each day, the tasks were performed in order 1 to 4. The volunteers were not allowed to switch between tasks freely. On the 1st day (Part #1) of the user study, a participant was not told why he was reading the materials. A participant was not told that he would make a presentation until the 19th day (Part #6). Before each presentation, a participant had several minutes to recall a search task using TimeTree or SearchLog. The search tasks were presented with TimeTree or SearchLog. During the presentations, the experts are free to ask questions.

The participants were asked to perform the search tasks in a controlled environment. The participants were told not to do anything related to the search tasks when they are not in the controlled environment. The participants were also told not to talk about the user study to each other.

5. Results analyses

The following abbreviations are used in this section: Avg. = Average, SD = Standard Deviation, TT = TimeTree, SL = SearchLog, TTy = Task Type, Cor = Correlation coefficient.

5.1. Subjective evaluation

We took 2 surveys in the user study. The first survey was taken on the 10th day (Part #5) of the user study. The first survey was about the subjective experience on TimeTree during the search process. The results are shown in Table 2. According to Question #1 & #2, most participants agreed that TimeTree was easy to learn and operate. According to Question #3, TimeTree did cause extra burden during the search process. During the interviews, the participants agreed on the main burden caused by TimeTree. The main burden was to drag new nodes that showed up in wrong locations to correct locations. The wrong locations of new nodes were caused by the TF/IDF based query provenance identification method. The TF/IDF based method might not be able to identify the correct source for a new query node. Thus, the query new node might be put to a wrong location. We may be able to overcome such a disadvantage by monitoring the document a user is reading. Meanwhile, according to Question #4, although with a larger standard deviation, the participants trended to thought that TimeTree brought more benefits than drawbacks.

Table 2
Subjective feedbacks taken on the 10th day (Part #5) of the user study. For score range 0/1: 0 is for no; 1 is for yes. For score range 0–10: 0 is for strongly disagree; 5 is for neutral; 10 is for strongly agree.

Question #	Question	Score range	Avg. score	SD
1	TimeTree was easy to learn.	0–10	9.281	0.729
2	TimeTree was easy to operate.	0–10	8.531	0.915
3	TimeTree caused no extra burden.	0–10	7.843	1.194
4	In contrast to the extract burdens, TimeTree brought more benefits.	0–10	9.219	1.431
5	I have used the location of tree nodes to determine the temporal order of nodes.	0/1	0.938	n/a
6	I have used the parent–child relation of tree nodes to remember why I searched for a query.	0/1	0.875	n/a
7	I have used the tree structure to identify sub-tasks.	0/1	0.875	n/a
8	Each time I needed to resume a type 1 task, TimeTree helped me remember what I had accomplished and what to do next.	0–10	8.969	1.062
9	TimeTree intuitively visualized my type 1 tasks.	0–10	8.656	0.865
10	Each time I needed to resume a type 2 task, TimeTree helped me remember what I had accomplished and what to do next.	0–10	9.156	0.723
11	TimeTree intuitively visualized my type 2 tasks.	0–10	9.031	0.767

Table 3
Subjective feedbacks taken on the 19th day (Part #6) of the user study. Score meanings are the same as Table 2.

Question #	Question	Score range	Avg. score (TT)	SD (TT)	Avg. score (SL)	SD (SL)	Paired t-test p-value
	The system helped me to remember:						
12	(1) the temporal order of queries and clicks;	0–10	8.5	0.894	7.813	1.602	**0.0769**
13	(2) the reason why I searched a query;	0–10	8.75	0.842	6.563	1.664	1.92e−4
14	(3) the sub-tasks.	0–10	9.063	0.840	6.281	2.655	5.82e−4

Table 4
Expert evaluations on the presentations with TimeTree and SearchLog.

Quality measure #	Quality measure on the presentations	Score range	Avg. score (TT)		Avg. score (SL)		Paired t-test p-value	
			TTy1	TTy2	TTy1	TTy2	TTy1	TTy2
1	Temporal coherence	0–10	7.953	8	8	7.984	**0.728**	**0.708**
2	Causal coherence	0–10	8.281	8.016	7.156	7.109	8.787e−6	1.397e−5
3	Thematic coherence	0–10	8.125	7.938	7.109	7.063	0.000117	1.349e−5
4	Overall quality	0–10	8.063	7.922	7.344	7.281	0.00840	0.00147

Question #5 to #7 were asked to see whether the participants were familiar with the RCST structure of TimeTree. The results showed that most participants had used RCST to extract temporal, causal and thematic information during the search process. Question #9 and #11 stated that most participants found TimeTree to be intuitive. Question #8 and #10 also stated that most participants thought that TimeTree helped them in resuming search tasks.

The second survey was taken on the 19th day (Part #6) of the user study. The second survey was about the subjective experience on TimeTree and SearchLog during the recollection process. The results are shown in Table 3. The results showed that TimeTree gained statistically significantly higher scores on Question #13 and #14. These scores meant that the participants thought TimeTree helped them to achieve better causal coherence and thematic coherence than SearchLog. While according to Question #12 we could see that both TimeTree and SearchLog performed well in supporting temporal coherence. These results showed that subjectively, TimeTree helped maintain all the 3 types of global coherence.

5.2. Expert evaluation

On the 19th day (Part #6) of the user study, two Ph.D. students were asked to work as experts to score the presentations. Scores were given in terms of 4 quality measures, as shown in Table 4. The results were consistent with the subjective evaluations. When using TimeTree, the participants gained significantly higher scores in terms of causal coherence, thematic coherence and overall quality. The two systems also gained similar scores on temporal coherence. The results showed that compared to SearchLog, TimeTree helped maintain all the 3 types of global coherence for complex search experiences. The results also showed that TimeTree helped the participants gain better presentation quality. Such improvements were also consistent with the research of [10].

5.3. Objective evaluation

5.3.1. Statistics on the search tasks

We counted the numbers of query operations and click operations in search tasks as Table 5. Given a query, requesting a new page of search results was considered as a separate query operation.

Table 5
Statistics on the numbers of query operations and click operations in search tasks.

	TTy1	TTy2
Avg. query ops	43.407	33.906
Avg. click ops	43.625	24.094
Paired t-test p-value	0.167	4.43e−6

Table 6
Participants' performance in maintaining causal coherence and thematic coherence in different systems.

Coherence	TT		SL		F-test p-value		t-test p-value	
	TTy1	TTy2	TTy1	TTy2	TTy1	TTy2	TTy1	TTy2
Causal coherence: average number of queries with clear reasons per task (out of 10 randomly chosen queries)	8.781	9.217	8.844	9.375	0.993	0.672	0.879	0.376
Thematic coherence: average number of sub-tasks summarized per task	4.438	2.563	4.313	2.375	0.288	0.382	0.708	0.606

We analyzed the numbers of query operations and click operations by paired Student's t-test. The analysis showed no significant difference between different systems and between tasks of the same type. However, the numbers of click operations differed significantly between different types of task. The difference meant that the participants clicked less results in type 2 tasks than in type 1 tasks. Such a behavioral difference meant that the participants formed different types of search process in different types of search task.

5.3.2. Participants' performance in maintaining causal coherence and thematic coherence

According to Sections 5.1 and 5.2, the participants and the experts felt that TimeTree and SearchLog were (statistically) significantly different in helping the participants to maintain causal coherence and thematic coherence. We wondered what caused the different feelings. To figure out the cause, we first studied the participants' performance in maintaining causal coherence and thematic coherence with TimeTree and SearchLog. On the 19th day (Part #6) of the user study, after the presentation of each search task, the participant was asked to summarize the sub-tasks of the search task. We also randomly chose 10 queries from the search task. We then asked the participant to explain why he searched for the queries. The average number of sub-tasks summarized per task, and the average number of queries with clear reasons are shown in Table 6. From the results, we could see that the participants summarized almost the same number of sub-tasks per task using the two systems. The participants also gave clear reasons to almost the same number of queries using the two systems. The results showed that, when focusing on a specific type of coherence, the performance of the participants was almost the same on the two systems.

5.3.3. Features benefiting/harming causal coherence and thematic coherence

The results of Section 5.3.2 seemed to be in conflict with the results of Sections 5.1 and 5.2. We wondered why participants with similar performance on TimeTree and SearchLog gained different scores when using different systems. To answer this question, we extracted features that may benefit, or harm causal coherence and thematic coherence of the presentations given by the participants. The features were extracted from the recording of the presentations of the participants. For causal coherence, we focused on why a participant issued a query. We extracted all the fragments related to the reasons of queries from the recordings. We then extracted features according to whether a feature could help clarify the reason of a query. The extracted features benefiting and harming causal coherence are listed as below:

- Features benefiting causal coherence
  - Clear source descriptions, e.g. I read about ... so I searched ...
- Features harming causal coherence
  - Lack of certainty in source descriptions: e.g. There should be a ... talking about ...
  - Lack of causal descriptions
  - Unable to answer causal questions asked during the presentation

For causal coherence, we focused on whether a participant could clearly specify the topic structure of a search task. We extracted all the fragments related to topic descriptions from the recordings. We then extracted features according to whether a feature could help clarify the topic structure of the search task. The extracted features benefiting and harming thematic coherence as listed as below:

- Features benefiting thematic coherence

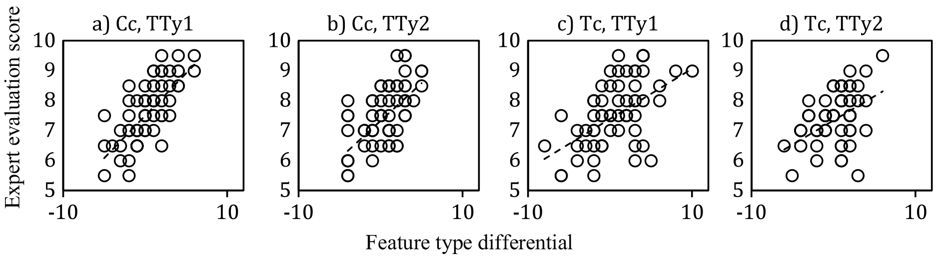


Fig. 5. Task distributions plotted in the space of feature type differential vs expert evaluation score. Each circle represents a search task. Feature type differentials are calculated by subtracting the numbers of harmful features from the numbers of beneficial features. Cc=Causal coherence, Tc=Temporal coherence.

Table 7
Results of the Pearson's correlation test between the feature type differentials and the expert evaluation scores.

Coherence	Cor	*p*-value
Causal coherence, TTy1	0.729	8.515e−12
Causal coherence, TTy2	0.608	9.981e−8
Thematic coherence, TTy1	0.616	0.000174
Thematic coherence, TTy2	0.495	0.00401

Table 8
Average feature type differentials on the two systems.

Coherence	Avg. differential		Paired *t*-test
	TT	SL	*p*-value
Causal coherence, TTy1	1.656	−0.969	1.033e−5
Causal coherence, TTy2	1.094	−0.188	0.0134
Thematic coherence, TTy1	1.625	−0.438	0.00783
Thematic coherence, TTy2	0.938	−0.563	0.000623

  - Clear topic descriptions for queries and clicks
- Features harming thematic coherence
  - Unclear topic descriptions: e.g. maybe..., sort of..., these are some related papers (without telling what the papers are related to)
  - Unable to answer topical questions asked during the presentation

We then counted the number of beneficial and harmful features for each task. We then calculated the feature type differential for each task by subtracting the number of harmful features from the number of beneficial features of both TimeTree and SearchLog. The results are shown in Fig. 5. From Fig. 5a to d we could see that the feature type differentials seemed to be linearly correlated to the expert evaluation scores. Pearson's correlation test was used to determine the correlation, as shown in Table 7. The results showed that, the feature type differentials were linearly correlated to the expert evaluation scores for causal coherence and for thematic coherence. The average feature type differentials for TimeTree and SearchLog are shown in Table 8. The results showed statistically significant differences between TimeTree and SearchLog for causal coherence and thematic coherence. The results also indicated that the participates showed higher feature type differentials in TimeTree than in SearchLog. We argue that such behavioral changes lead to the improved scores of causal coherence and thematic coherence in the subjective evaluations and the expert evaluations.

5.3.4. User behavioral patterns in TimeTree

We wondered why the participants showed higher feature type differentials in TimeTree. To understand the reason, we studied how the participants used TimeTree when recollecting past search tasks. We wanted to know how the participants rebuilt the causal and the topical relations between queries and clicks. Thus, we analyzed the hops between the adjacently gazed tree nodes, as shown in Fig. 6. The eye gaze data was collected with a Tobii eye tracker when the participants presenting their search processes. The results showed that 86.3% of the adjacency gazes happened on nodes that are within 3 hops from each other. The locational patterns of the adjacently gazed nodes are shown in Table 9.

The results showed that most adjacent gazes happened between sibling nodes and between parent–child nodes. The topics of sibling nodes, as well as parent–child nodes, are usually closely related. These topically related nodes could be used by the participants to conclude sub-tasks from past search tasks, thus maintaining thematic coherence.

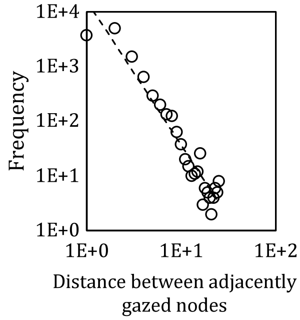
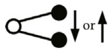
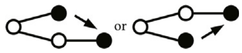
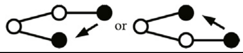


Fig. 6. Distribution of the hops between adjacently gazed tree nodes.

Table 9
Locational patterns of the adjacently gazed tree nodes. Black discs indicate gazed nodes. Arrows indicate the sequence of two adjacently gazed nodes.

Hop	% of all gazes	Pattern	% of hop i gazes
1	31.4		46.5
			53.5
2	42.1		93.7
3	12.8		45.9
			50.2

We also noticed another interesting result in Table 9. For adjacently gazed nodes that are parent–child nodes (hop-1 gazes), the child-to-parent pattern occurred even more often than the parent-to-child pattern. This meant that the participants used the source-tracking feature a lot to identify the sources of queries and clicks. We think this fact led to the improvements in causal coherence when using TimeTree.

6. Conclusion and discussion

In this paper, we introduced our coherence-oriented complex search experience management method TimeTree to maintain temporal, causal and thematic coherence for complex search experiences. By conducting a user study, we learned how and to what extent TimeTree helped the participants maintain temporal coherence, causal coherence, and thematic coherence. We discuss some of the questions emerged during this work.

The subjective evaluation results showed that TimeTree may cause extra burden. The user interviews showed that the extra burden was caused by dragging a new query node to its correct location when TimeTree failed to identify the correct source for the new query. Since the user study relied on a TF/IDF based source-tracking method to form the tree structure of the RCST, more sophisticate methods should be studied to identify the source of a query.

The objective evaluation results in Section 5.3.2 showed that the performance of the participants was almost the same on the two compared systems when focusing on a specific type of coherence. However, in Section 5.3.3, we did observe behavioral changes that may lead to the improved scores in maintaining thematic coherence and causal coherence. We thus infer that TimeTree did not improve the participants' abilities in maintaining different types of coherence for complex search experiences. Instead, TimeTree made the experiences about complex search tasks more accessible to improve the overall performance of presenting past search tasks. We believe such a finding may help the design of new methods in managing complex search experiences, and even in supporting search task resumption and search result synthesizing in multi-session complex search.

The 2 hop-3 patterns shown in Table 9 seemed to be interesting. However, further investigation showed that most of the adjacent gazes conforming to the 2 patterns happened on the nodes that were located beside each other. This meant that the participants might just be gazing at a node, and inadvertently noticed another node located beside the node. We currently do not know whether these nodes that were located beside a node by chance affect the recollections of past search tasks. But the 2 behavioral patterns reminded us to pay more attention on the layout design of new search task visualization methods.

Acknowledgments

This research is supported by the National Natural Science Foundation Program of China (61502089, 61603082, 61572116, 61572117), the National Key Technology Support Program, China (2014BAI17B00), the National Key Technology R&D Program of the Ministry of Science and Technology, China (2015BAH09F02 2015BAH47F03), and the Special Fund for Fundamental Research of Central Universities of Northeastern University, China (N161604009).

References

[1] Dan Morris, Meredith Ringel Morris, Gina Venolia, Searchbar: a search-centric web history for task resumption and information re-finding, in: CHI, 2008, pp. 1207–1216.
[2] L. Helen Williams, A. Margin Conway, Gillian Coben, Autobiograpical memory, in: Memory in the Real World.
[3] Jiafeng Guo, Xiaofei Zhu, Yanyan Lan, Xueqi Cheng, Modeling users' search sessions for high utility query recommendation, Inf. Retrieval J. 20 (1) (2017) 4–24.
[4] Huanhuan Cao, Daxin Jiang, Jian Pei, Qi He, Zhen Liao, Enhong Chen, Hang Li, Context-aware query suggestion by mining click-through and session data, in: SIGKDD, 2008, pp. 875–883.
[5] Salvatore Andolina, Khalil Klouche, Jaakko Peltonen, et al., Intentstreams: Smart parallel search streams for branching exploratory search, in: IUI, 2015, pp. 300–305.
[6] Wei-Lin Chen, Wei-Guang Teng, Exploiting browsing history for exploratory search, in: HCI, 2009, pp. 355–364.
[7] Hogun Park, Sung-Hyon Myaeng, Gwan Jang, Jong-wook Choi, Sooran Jo, Hyung-chul Roh, An interactive information seeking interface for exploratory search, in: ICEIS, 2008, pp. 276–285.
[8] Tuukka Ruotsalo, Giulio Jacucci, Petri Myllymäki, Samuel Kaski, Interactive intent modeling: information discovery beyond search, Commun. ACM 58 (1) (2015) 86–92.
[9] Bahareh Sarrafzadeh, Olga Vechtomova, Vlado Jokic, Exploring knowledge graphs for exploratory search, in: IIiX, 2014, pp. 135–144.
[10] Tilmann Habermas, Susan Bluck, Getting a life: The emergence of the life story in adolescence, Psychol. Bull. 126 (5) (2000) 748–769.
[11] Rosie Jones, Kristina Lisa Klinkner, Beyond the session timeout: Automatic hierarchical segmentation of search topics in query logs, in: CIKM, 2008, pp. 699–708.
[12] Susan, J. Feinglos, MEDLINE at BRS, DIALOG, and NLM: is there a choice?, Bull. Med. Libr. Assoc. 71 (1) (1983) 6–12.
[13] Georg Singer, Dmitri Danilov, Ulrich Norbisrath, Complex search: Aggregation, discovery, and synthesis, Proc. Estonian Acad. Sci. 61 (2) (2012) 89–106.
[14] Tingting Jiang, Exploratory search: A critical analysis of the theoretical foundations, system features, and research trends, in: Library and Information Sciences, 2014.
[15] Nicholas J. Belkin, Charles L.A. Clarke, Ning Gao, Jaap Kamps, Jussi Karlgren, Report on the SIGIR workshop on "entertain me": Supporting complex search tasks, ACM SIGIR Forum 45 (2) (2011) 51–59.
[16] Maria Gäde, Mark M. Hall, Hugo C. Huurdeman, et al., Report on the first workshop on supporting complex search tasks, ACM SIGIR Forum 49 (1) (2015) 50–56.
[17] Georg Singer, Ulrich Norbisrath, Dirk Lewandowski, Ordinary search engine users assessing difficulty, effort, and outcome for simple and complex search tasks, J. Inf. Sci. 39 (3) (2013) 346–358.
[18] Marc-Allen Cartright, Ryen W. White, Eric Horvitz, Intentions and attention in exploratory health search, in: SIGIR, 2011, pp. 65–74.
[19] Eric Nyberg, Yi Yang, Leveraging procedural knowledge for task-oriented search, in: SIGIR, 2015, pp. 513–522.
[20] Ahmed Hassan Awadallah, Ryen W. White, Patrick Pantel, Susan T. Dumais, Yi-Min Wang, Supporting complex search tasks, in: CIKM, 2014, pp. 829–838.
[21] Jaspreet Singh, Wolfgang Nejdl, Avishek Anand, Expedition: A time-aware exploratory search system designed for scholars, in: SIGIR, 2016, pp. 1105–1108.
[22] Alan Medlar, Kalle Ilves, Ping Wang, Wray, L. Buntine, Dorota Glowacka, PULP: A system for exploratory search of scientific literature, in: SIGIR, 2016, pp. 1133–1136.
[23] Phong H. Nguyen, Kai Xu, Ashley Wheat, B.L. William Wong, Simon Attfield, Bob Fields, Sensepath: understanding the sensemaking process through analytic provenance, IEEE Trans. Vis. Comput. Graphics 22 (1) (2016) 41–50.
[24] Katrin Lamm, Thomas Mandl, Ralph Koelle, Search path visualization and session performance evaluation with log files, in: CLEF, 2009, pp. 538–543.
[25] Marcus Nitsche, Andreas Nürnberger, Trailblazing information: an exploratory search user interface, in: HCI, 2013, pp. 230–239.
[26] Phong H. Nguyen, Kai Xu, Andy Bardill, Betul Salman, Kate Herd, B. L. William Wong, Sensemap: Supporting browser-based online sensemaking through analytic provenance, in: VAST, 2016, p. 91-100.
[27] Xiaolong Zhang, Yan Qu, C. Lee Giles, Piyou Song, Citesense: supporting sensemaking of research literature, in: CHI, 2008, pp. 677–680.
[28] Justin Matejka, Tovi Grossman, George, W. Fitzmaurice, Citeology: visualizing paper genealogy, in: CHI Extended Abstracts, 2012, pp. 181–190.
[29] Duen Horng Chao, Aniket Kittur, Jason I. Hong, Christos Faloutsos, Apolo: making sense of large network data by combining rich user interaction and machine learning, in: CHI, 2011, pp. 167–176.
[30] Susan Bluck, Autobiographical memory: Exploring its functions in everyday life, Memory 11 (2003) 113–123.
[31] Elise van den Hoven, Berry Eggen, The design of a recollection supporting device - A Study into triggering personal recollections, in: HCI, 2003, pp. 1034–1038.
[32] Harumi Murakami, Kenta Mitsuhashi, Keita Senba, Creating user's knowledge space from various information usages to support human recollection, in: IEA, 2012, pp. 596–605.
[33] Zhi Sun, Qinke Peng, Jia Lv, Tao Zhong, Analyzing the posting behaviors in news forums with incremental inter-event time, Physica A 479 (2017) 203–212.
[34] Abigail Sellen, Andrew Fogg, Mike Aitken, Steve Hodges, Carsten Rother, Kenneth, R. Wood, Do life-logging technologies support memory for the past: an experimental study using sensecam, in: CHI, 2007, pp. 81–90.
[35] Shoko Kodama, Hideo Akaike, Hiroyasu Kakuda, Lifelog sharing system for memory recollection, in: GCCE, 2014, pp. 81–83.
[36] John, H. Mace, The Act of Remembering: Toward an Understanding of how We Recall the Past, Wiley-Blackwell, 2010.
[37] Annie Piolat, Thierry Olive, Ronald, T. Kellogg, Cognitive effort during note taking, Appl. Cognit. Psychol. 19 (2005) 291–312.
[38] Ryen W. White, Bill Kules, Steven M. Drucker, M.C. schraefel, Supporting exploratory search, Commun. ACM 49 (4) (2006) 37–39.