PaperPoles: Facilitating Adaptive Visual Exploration of Scientific Publications by Citation Links[1]

Jiangen He* and Qing Ping
Department of Information Science, College of Computing and Informatics, Drexel University, Philadelphia, USA. E-mail: jiangen.he@drexel.edu; qp27@drexel.edu

Wen Lou
Department of Information Management, Faculty of Economics and Management, East China Normal University, Shanghai, China. E-mail: wlou@infor.ecnu.edu.cn

Chaomei Chen
Department of Information Science, College of Computing and Informatics, Drexel University, Philadelphia, USA. E-mail: chaomei.chen@drexel.edu

Abstract

Finding relevant publications is a common task. Typically, a researcher browses through a list of publications and traces additional relevant publications. When relevant publications are identified, the list may be expanded by the citation links of the relevant publications. The information needs of researchers may change as they go through such iterative processes. The exploration process quickly becomes cumbersome as the list expands. Most existing academic search systems tend to be limited in terms of the extent to which searchers can adapt their search as they proceed. In this paper, we introduce an adaptive visual exploration system named PaperPoles to support exploration of scientific publications in a context-aware environment. Searchers can express their information needs by intuitively formulating positive and negative queries. Search results are grouped and displayed in a cluster view, which shows aspects and relevance patterns of the results to support navigation and exploration. We conducted an experiment to compare PaperPoles with a list-based interface in performing two academic search tasks with different complexity. The results show that PaperPoles can improve the accuracy of searching for the simple and complex tasks. It can also reduce completion time of searching and exploration effectiveness in the complex task. PaperPoles demonstrates a potentially effective workflow for adaptive visual search of complex information.

[1] This a preprint of an article accepted for publication in JASIST - Journal of the Association for Information Science and Technology (October 28, 2018).

* corresponding author.

Introduction

Finding relevant publications is an essential first step in a research process. One common method of finding publications is citation chaining (also known as snowballing) in which researchers follow citation links between publications (Athukorala, Hoggan, Lehtiö, Ruotsalo, & Jacucci, 2013; Bates, 1989; Ellis, 1989; Ellis, Cox, & Hall, 1993). Based on a publication identified from a search, searchers can follow either the publications that cited it (forward chaining) or the publications in its reference list (backward chaining). The chaining activities may simulate serendipitous information discovery and even research creativity, as both backward and forward citations may not be on the same topic of the identified publications in the conventional sense. Because of the diverse topics of citations, researchers may expand the search indefinitely in any direction (Bates, 1989). The exploratory features of chaining activities are desirable for information-seeking behaviors that support problem-solving and opportunity-finding process in scientific research. However, it is challenging to explore the citations over traditional academic search systems because of their strong dependency on precise searcher-generated queries and list-based interfaces (di Sciascio, Sabol, & Veas, 2016).

Academic searchers may have complicated, uncertain, and evolving information needs when exploring citations, which make formulating keyword queries difficult in many scenarios (Li, Schijvenaars, & de Rijke, 2017). For example, a searcher may like to explore an intersection of two or more research paths respectively represented by different identified publications, which may be research opportunities where novelty is high, and a searcher may change their interest because of a serendipitous discovery. The learning and investigative process (Marchionini, 2006) may require contextual information to help searchers define and express their information needs.

List-based views in a traditional academic system display results in a relevance-based order, but they may not support showing navigational guidance or allow multiple iterations. Automatically grouping the citations according to their common features can reduce cognitive load and facilitate activities of learning and drilling down. Although the list-based views can show the grouping information by augmented lists (Iwata et al., 2012) or multiple lists (Görg et al., 2013), they scarcely show overviews of semantic groups, which can reveal the differences among the groups.

Current methods to support the exploratory search improved the engagement of searchers control over the search process by showing multiple aspects and/or relevance distribution of search results and modeling information needs of the searchers (Ahn & Brusilovsky, 2013; Clarkson, Desai, & Foley, 2009; Felix, Pandey, & Bertini, 2017; Klouche, Ruotsalo, Micallef, Andolina, & Jacucci, 2017; Peltonen, Belorustceva, & Ruotsalo, 2017; Ruotsalo et al., 2013; Smith et al., 2006). Some methods were designed to support the exploratory search in academic contexts (Athukorala et al., 2017; Mackinlay, Rao, & Card, 1995; Ponsard, Escalona, & Munzner, 2016). Despite existing evidence on the critical role of chaining activities in academic information-seeking behavior (Barrett, 2005; Bates, 1989; Ellis, 1989; Ellis et al., 1993; Meho & Tibbo, 2003; Rhee, 2012; Savolainen, 2017b), there are few methods available to support exploration of publications in an interactive user interface with an aspect and relevance overview of search results, in which searchers can explore publications by making use of citation links instead of text queries.

We propose an interactive visualization system named PaperPoles for exploring citations of publications. The "Pole" of PaperPoles is based on a metaphor of the two poles of a magnet. A magnet is composed of a south pole and a north pole which have opposite attraction to other magnets. In analogy, a positive "pole" and a negative "pole" are designed in the system to enable searchers to express their wanted and unwanted information intuitively. Based on the information needs of the searchers, PaperPoles automatically groups the citations into clusters, labeling the clusters with text analytics techniques, and visualize the clusters and relevance information of citations. The visualization can guide searchers to get into the approximately right subset of citations. The searchers can also refine their queries iteratively and access the updated result immediately. Multiple coordinated views provide different levels of detail of the citation set and allow the searchers to interactively drill down into subsets in the context of a cluster overview and their relevance patterns.

The interface was evaluated in a user study with 28 participants. The participants performed two search tasks with different complexity on PaperPoles and a baseline system. They were asked to find publications which are relevant to a research topic and the tasks started with the publications they were given. Our results show PaperPoles significantly improved the effectiveness measured by accuracy in both simple and complex tasks, but PaperPoles improved the efficiency measured by completion time and the quality of viewed information only in the complex task. We also analyzed the action usage of participants for completing the tasks on the two systems.

Our contributions are: (1). We propose a visual analytics system for seamlessly creating and refining complex positive and negative queries and a visualization showing both an aspect overview and relevance patterns of search results. (2). We provide evidence that the visualization and interaction designed in the system can help searchers in exploratory search tasks, especially in complex tasks.

Related Work

Exploratory Search and Academic Search

Berrypicking model can be regarded as a prototypical approach to exploratory search (Bates, 1989; Savolainen, 2017a). Using the metaphor of picking huckleberries from a bush, the berrypicking model depicts information searching as an evolving activity during which searchers pick up pieces of information as they navigate through an information space. In this evolving process, searchers acquire new knowledge and change their perception of search tasks through their interaction with the information space (Belkin et al., 1993). Thus, the search process is considered to be cognitively complex with the information seeker being uncertain about the search and the task goals are usually imprecise and open-ended (Athukorala, Głowacka, Jacucci, Oulasvirta, & Vreeken, 2016; Kim, 2009; White & Roth, 2009). A common exploratory search strategy observed by many studies is: a searcher starts with a tentative query to navigate to the approximately right part of the information space, and then leverage contextual cues to choose subsequent exploration steps until search goals are achieved (Kairam, Riche, Drucker, Fernandez, & Heer, 2015; O'Day & Jeffries, 1993; Ruotsalo et al., 2013; Teevan, Alvarado, Ackerman, & Karger, 2004).

Academic search is usually complex and exploratory in nature because of the complexity of academic tasks (Du & Evans, 2011). Exploratory search is a common approach to dealing with the complexity of academic search (T. Chen & Gross, 2017). Therefore, it is reasonable to see the exploratory search tasks in many studies were situated within an academic context (Athukorala et al., 2016; Bates, 1989). Du and Evans (2011) identified the search strategies developed by academic searchers, such as interaction with multiple systems, multiple queries, and reformulation, which are in line with characteristics of exploratory search. Exploration was found to be one of the most frequent purposes of academic search, but supporting exploration in academic search is not properly addressed and academic search tools should include techniques to support exploratory search (Athukorala et al., 2013).

Information-seeking behaviors in academic tasks have been studied extensively. Ellis and his associates identified six generic characteristics of information-seeking behaviors among physical and social scientists: starting, chaining, browsing, differentiating, monitoring, and extracting (Ellis, 1989; Ellis et al., 1993). Ellis's seminal models were elaborated by many empirical studies where novel components were added by redefining or by restructuring the components of the framework (Savolainen, 2017b). Meho and Tibbo (2003) identified another four behavior patterns of social scientists, namely, accessing, networking, verifying, and information management. Barrett (2005) found the information-seeking behaviors of graduate students in the humanities involved browsing, citation chasing, and constantly reading. Selecting Meho and Tibbo's model as a point of departure, Rhee (2012) conceptualized the patterns of historians' information-seeking behavior within diverse stages of the information-seeking process: searching, assessing, processing, and ending stages. These studies have revealed the importance of chaining (i.e. citation links) in academic information-seeking tasks.

Literature Collection Visualization

Many visualization methods and tools are available to support the analysis and exploration of a scientific literature collection (Federico, Heimerl, Koch, & Miksch, 2017). Although these methods and tools may not be designed specifically for search tasks, the general goal of their designs is to visualize the underlying structure of literature collection, which can also serve for the goals of search tasks.

Some of these tools (C. Chen, 2006; Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2012; Heimerl, Han, Koch, & Ertl, 2016; van Eck & Waltman, 2010) can assist us in learning scientific patterns and emerging trends. They show scientific structure and evolution by grouping publication, authors, institutes, words or other entities. Citation-based analysis is widely used by these visualization methods or tools. Based on the patterns identified by citation-based analysis and other analytic techniques, they provide overviews of scientific development with critical information. Some visualization systems were designed for exploring a literature collection and identifying papers of interest, but they can also visualize the underlying structure of a paper collection for the exploration (Beck, Koch, & Weiskopf, 2016; Berger, McDonough, & Seversky, 2017; Dunne, Shneiderman, Gove, Klavans, & Dorr, 2012; Lee, Czerwinski, Robertson, & Bederson, 2005). For example, Action Science Explorer (Dunne et al., 2012) constructs a citation network visualization and cite2vec (Berger et al., 2017) provides a word-document 2D projection view.

The visualization of a literature collection serves a proxy to build an overview understanding of the collection that can be helpful for various search tasks, especially for complex ones, because insights learned from the overview can facilitate the formulation of a high-quality query and direct exploration for complex information needs (White & Roth, 2009).

Search Result Visualization

Existing studies applied information visualization and information processing methods to help searchers learn and manipulate search results. We categorize these visualization interfaces into list-based search result visualization and non-list-based search result visualization.

List-based visualization interfaces display search results by a traditional list of items, but with visual augmentation to display an overview of result items. TileBar (Hearst, 1995), HotMap (Hoeber & Yang, 2006), and AspecTiles (Iwata et al., 2012) have similar visual designs using a tile mosaic metaphor. They display an overview of a search result by a combination of squares in a rectangle. uRank (di Sciascio et al., 2016) presents a result list with augmented document titles and stacked bars indicating relevance scores to improve transparency of relevance computation. Khazaei and Hoeber (2017) also visualized citations to support academic search as us, but they used a dual list view augmented by a glyph visualization to provide additional information. Tsai and Brusilovsky (2018) introduced a visual diversity-enhanced list view with other visualization methods for showing recommendations. Some interfaces assist searchers to formulate queries in list-based interfaces by supporting visual queries (Zhu & Yan, 2016) or providing a visual query suggestion interface enhanced (Umemoto, Yamamoto, & Tanaka, 2016). These list-based visualizations can be naturally incorporated with traditional list representations of IR systems, thus they can keep a clear view for determining the relevance of individual document as traditional list views. However, these visualizations only depict overviews of individual documents and do not provide an overview of the entire retrieved documents.

Non-list-based search result visualizations present overviews of result sets to relieve the browsing burden on searchers. Their visual designs tend to be more diverse than list-based ones. FacetMap (Smith et al., 2006) extracts facets of results from metadata and shows the facets that scaled naturally with the size of the dataset. ResultsMaps (Clarkson et al., 2009) uses hierarchical treemap representations with query string-driven digital library search engines. WebSearchViz (Nguyen & Zhang, 2006) and Adaptive VIBE (Ahn & Brusilovsky, 2013) both build a visual space to illustrate the relevance degrees among queries and search results. They also allow searchers to manipulate key visual elements to adapt the visualization. TextTiles (Felix et al., 2017) gives a dashboard with multiple coordinated views to enable an open-ended investigation of a text collection. Topic-relevance map (Peltonen et al., 2017; Ruotsalo et al., 2013) visualizes a topical overview of search results where the layout of keywords is used to show relevance and topical similarity. Most of the current methods focus on text-based points of interest, such as keywords or topics, to show the overview and direct the exploration. The links between documents, on which a common information-seeking behavior (chaining) is based (Kim, 2009), were not fully utilized.

Beyond displaying search results, many visualization systems aim to support iterative search process. They provide a highly interactive interface to engage searchers more fully in the search process and put them in continuous control (Marchionini, 2006), where both learning and investigating activities are supported. SenseMaker (Baldonado & Winograd, 1997) supports the evolution of a user's interest by an interface showing contextual information and expansion actions. Apolo (Chau, Kittur, Hong, & Faloutsos, 2011) guides the user to incrementally and interactively explore large network data by engaging users in bottom-up sensemaking. PivotPath (Dork, Riche, Ramos, & Dumais, 2012) and PivotSlice (Zhao, Collins, Chevalier, & Balakrishnan, 2013) provide an interface for faceted exploration of document relations and creating dynamic queries in a sensemaking process. VisIRR (Choo et al., 2018) dynamically and explicitly displays recommendations in the visualization to help searchers identify documents of interest and expand their search scope based on search preference.

System Design

PaperPoles [1] (see Figure 1) is a visualization system designed to assist searchers in exploring scientific publications by citation links. Initial queries are one or more seed papers, and results are citations of these seed papers. The searchers can interactively manipulate the results in a context-awareness environment to make the results meet their evolving information needs.

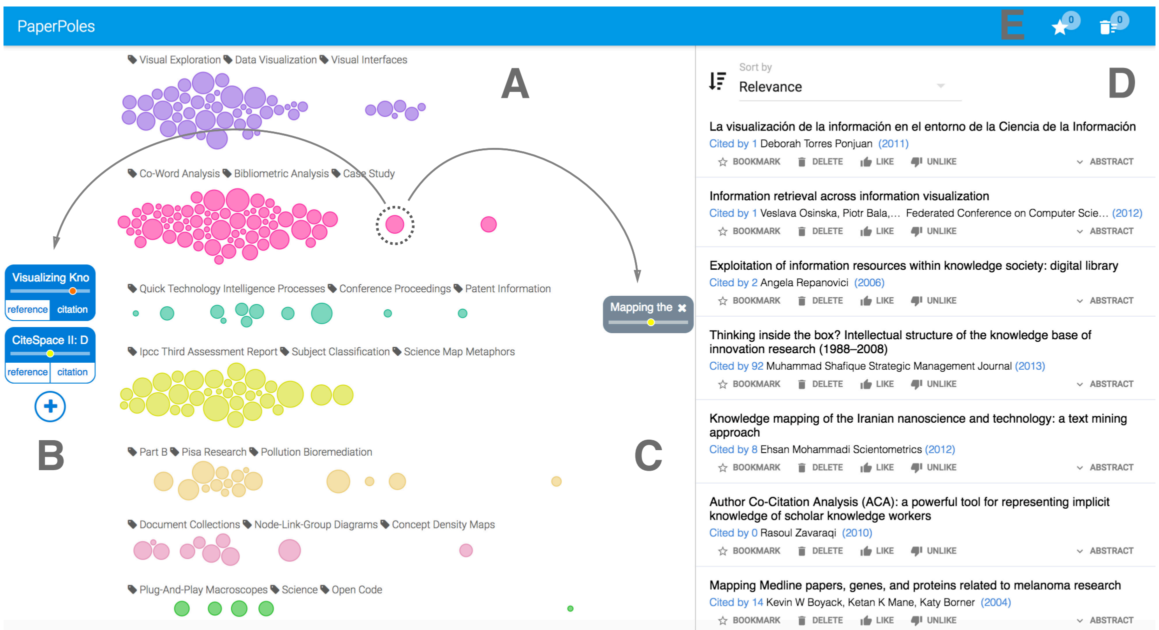

Figure 1. An overview of the proposed visualization system, PaperPoles, showing an example with two seed papers and one negative paper. PaperPoles contains four views: (A) the cluster view shows clusters of papers with labels; (B,C) the query boxes allow searchers to drag and drop positive and negative queries; (D) list view shows details of all the papers or selected ones and articles can be bookmarked, deleted, added as seed paper, or added as negative paper in the view; (E) the buttons are for viewing a list of bookmarked and deleted papers.

System Workflow

PaperPoles combines text analytics, a rank-augmented cluster view and flexible interaction among multiple coordinated views to assist in an exploratory search of scientific papers. Figure 2 depicts the workflow between automatic and interactive mechanisms. The workflow is summarized as follows:

[1] Web-based PaperPoles can be accessed at http://jiangenhe.com/paperpoles/.

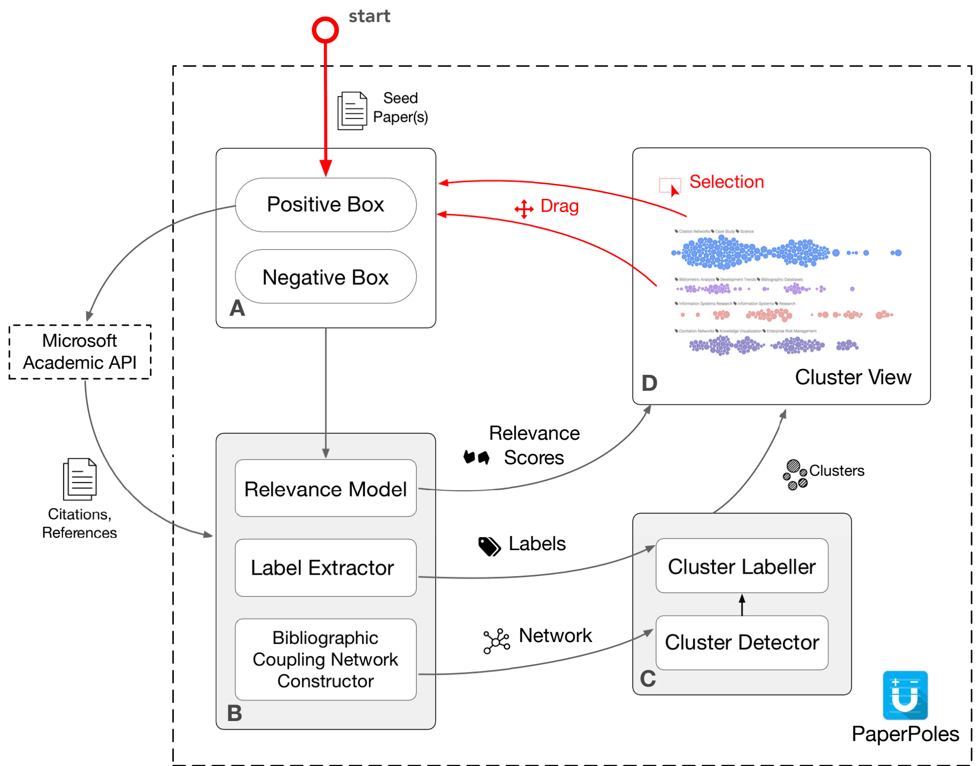

Figure 2. PaperPoles visual analytics workflow showing user-driven actions (red arrows) and automatic process (black arrows) of the system.

1. A searcher provides one or more identified relevant papers as seed papers to initialize the exploration (Figure 1.B and Figure 2.A).
2. The searcher decides to load the references or/and citations of the seed papers.
3. PaperPoles analyzes the loaded papers through three modules (Figure 2.B and C).
   - Relevance model computes relevance scores of papers to positive and negative queries.
   - Label extractor analyzes text (titles and abstracts).
   - Network constructor builds a bibliographic coupling network and detects clusters.
4. The interface displays a cluster visualization with labels (Figure 1.A and Figure 2.D) and a list view of papers (Figure 1.D).
5. By observing the clusters, the searcher can learn the underlying knowledge structure and relevance pattern of the papers.
6. The searcher interactively selects subsets for further investigation.
7. The searcher can add a paper into the positive and negative query boxes and tune their weights to express evolving information needs (Figure 1.B and C, and Figure 2.A).

User-driven actions (1, 5, 6 and 7) and automatic processes (2, 3 and 4) of the system form an iterative loop workflow.

Data Analysis Techniques

In this section, we describe our data analysis techniques that are mainly based on techniques of text analysis, network clustering and relevance model of documents.

Phrases Extraction. Noun phrases are extracted from both titles and abstracts of papers. To extract noun phrases, we perform a part-of-speech tagging (Toutanova & Manning, 2000) to identify nouns. To facilitate the relevance computation, we apply stemming algorithm to improve matching of words with the same word stem. Meanwhile, we produce a weight value for each phrase in a paper by computing TF-IDF (term frequency-inverse document frequency) (Jones, 1972).

Bibliographic Coupling Network. To reveal the underlying knowledge structure of a paper set, we construct a bibliographic coupling (Kessler, 1963) network of papers which is a commonly used method in science mapping (Boyack & Klavans, 2010). Specifically, similarities between papers i and j are decided by the references of i and j, and measured by Jaccard Index. If A is the set of papers that are cited by i, B is the set of papers that are cited by j and w is the weight of the link between i and j, then $w = \frac{|A \cap B|}{|A \cup B|}$. The low cost of data acquisition and processing can facilitate quick responses to user interaction.

Clustering and Labeling. We adopt Louvain method to partition the bibliographic coupling network into clusters, which is a greedy optimization method to reach a maximum of modularity of a partition of the network (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008). The high efficiency of this algorithm is helpful for quickly responding to the user-driven changes in the search results. However, the result of clustering usually includes clusters with very few papers because a small part of papers have no or very few bibliographic coupling relations with other papers. The small clusters may cause the waste of visual space if we visualize all the clusters. Therefore, we group papers from all of the clusters with papers less than five into a cluster named “uncategorized”.

We apply an automatic cluster labeling method to help searchers obtain a quick understanding of clusters. Candidates for a cluster label are noun phrases extracted from papers of this cluster. The phrases are ranked by TF-IDF weight.

Matching Clusters Over Time. Searchers can continuously load new papers to explore in PaperPoles, which may cause dramatic changes of clusters. The dramatic changes of the cluster view may confuse searchers and increase their mental demand for learning the new visualization. Nevertheless, some of the newly generated clusters are very similar to some of the previous clusters because many of papers remain in the same cluster after adding/removing a seed paper. Making a cluster inherit the order and color of a previous similar cluster can reduce the visual changes. To achieve this goal, we need to match clusters generated before and after the change of the results.

We applied a cluster matching method introduced by (Palla, Barabási, & Vicsek, 2007) and modified it according to our goal. The basic idea of the cluster matching method is to build a joint graph consisting of the union of links from the network before the change and the one after the change. Based on the clusters of the joint graph, we can investigate the time dependence of the clusters after the change to the clusters before the change. For example, if all the papers of a cluster before the change and a cluster after the change can be found in a joint cluster, the cluster after the change is one of the successors of the cluster before the change.

We denote the set of clusters before a change by **D**, the set of clusters after the change by **E**, and the set of clusters from the joint network by **V**. For any cluster $D_i \in \mathbf{D}$ or $E_j \in \mathbf{E}$ we can find one cluster $V_k \in \mathbf{V}$ containing it.

After building connection between cluster sets before and after the change through clusters of a joint network, we need to match the clusters $D_i^k \in \mathbf{D}$ and $E_j^k \in \mathbf{E}$ that are contained in V_k which means $D_i^k \subseteq V_k$ and $E_j^k \subseteq V_k$. Then the relative overlap between every possible (D_i^k, E_j^k) pair is defined as below:

$$C_{ij}^k = \frac{D_i^k \cap E_j^k}{D_i^k \cup E_j^k}$$

We match the pairs of clusters in descending order of their relative overlaps. If the number of D^k is larger than the number of E^k, we find the successor(s) for each D_i^k from E^k. Conversely, if the number of E^k is larger than the number of D^k, we find predecessor(s) of E_j^k from D^k. We do not display the splitting and merging of clusters over time. Therefore, if a cluster of D has multiple successors, only the successor with the largest relative overlap with the cluster inherits the visual property of the cluster and other clusters appears as new clusters; if a cluster of E has multiple predecessors, the cluster inherit the visual property of the predecessor with the largest relative overlap.

Relevance Computation. PaperPoles allows searchers to refine the relevance model by adding papers as positive and negative queries. The relevance scores r of a paper p to a paper query q with weight w_q are calculated as follows:

$$r(p, q) = \text{Jaccard}(\text{refs}(p), \text{refs}(q)) \times w_q,$$

where refs(p) is the set of references of p.

Given the relevance score is computed for multiple queries, we normalize the relevance scores via SUM norm algorithm (Montague & Aslam, 2001) and then combine the normalized scores by CombMNZ method (Fox & Shaw, 1994) to optimize the performance of the combination of multiple scores. The SUM norm is a shift and scale invariant that is fairly outlier insensitive. It shifts min to 0 and scales sum to 1. CombMNZ combines multiple scores as follows:

$$\text{CombMNZ}(r\text{s}) = \text{CombSUM}(r\text{s}) * num\_nonzero\_r\text{s},$$

where rs stands for relevance scores, and CombSUM is the sum of individual relevance score. The combination of SUM norm and CombMNZ performs well and stable on combining multiple relevance scores (Montague & Aslam, 2001).

After combining the normalized relevance scores for queries and negative queries respectively, we calculate the relevance score of a paper p for the set of positive queries $Q^{pos} = \{q_i^{pos}\}$ and the set of negative queries $Q^{neg} = \{q_i^{neg}\}$ as below.

$$r(p) = \sum r(p, q_i^{pos}) - \sum r(p, q_i^{neg})$$

Visual Interface

PaperPoles's interface (Figure 1) is arranged as multiple coordinated views that display different levels of details of a paper set:

Cluster view. The cluster view (Figure 1.A) shows underlying topics of the paper set. Each cluster presents a topic among the set and each cluster consists of paper nodes in this topic. The layout of nodes in a cluster is optimized by Force-Directed Layout algorithm to remove the overlaps between the nodes. The horizontal position of a paper node displays a combined relevance score of the paper to both positive and negative queries (Figure 3). The size of a node indicates the citation count of the corresponding paper. Three labels are selected to summarize the textual information of a cluster. The cluster view supports pan via mouse wheel and selection starting from white space. Hovering over a node, searchers can check the details of the paper in a popup tooltip. To show the nodes viewed by hovering, the saturation of viewed nodes would be lower.

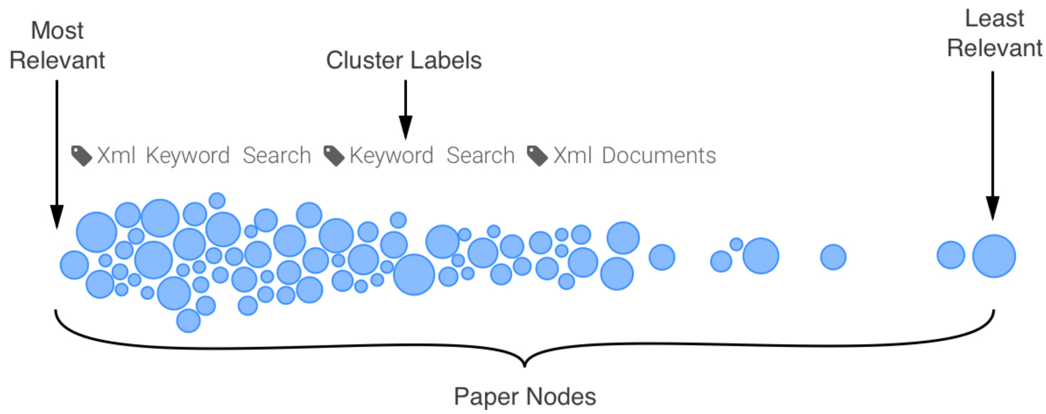

Figure 3. Cluster design.

Query drop box. A positive query drop box (Figure 1.B) and a negative query drop box (Figure 1.C) are placed on the two sides of the cluster view, into which a searcher drag paper nodes to refine queries. The searcher can also add a new seed paper to positive zone by clicking on the adding button (+) to trigger a search dialogue.

List view. The list view (Figure 1.F) displays the titles and other metadata of papers. When a subset of papers is selected in the cluster view, the list view only shows selected ones; when none of the paper is selected, the list view shows all the papers. The items on the list can be sorted by relevance score, the number of citations, and publication year. A paper can be added to positive and negative query boxes, bookmarked, and deleted on the list view. When a paper on the list is being hovered, its corresponding node on the cluster view would be highlighted by a red ring.

Interactions and Visual Design

Exploring Citations. We designed several interactions to help searchers explore topics of interest and early prune irrelevant topics.

Selection. Initially, the cluster view gives an overview of topics and their relevance pattern. PaperPoles allows searchers to select any subset of search results by drawing a rectangle in the cluster view (Figure 4), for example, searchers can select clusters or subsets with relatively higher relevance. The list view also only shows the selected papers. Searchers can further inspect if the selected subset is the one of their interest.

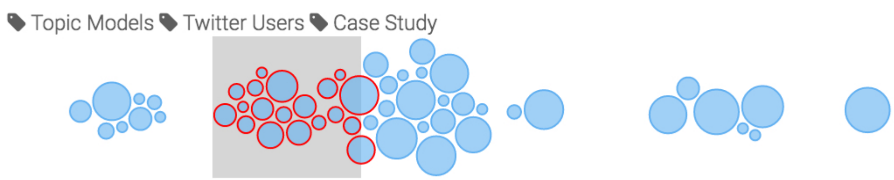

Figure 4. Selecting paper nodes for drilling down subsets of papers.

Deletion. During exploration, searchers may identify some irrelevant clusters or papers they do not want to see in their further exploration. The searchers can delete identified irrelevant papers in PaperPoles. A deleted paper won't be included in the set in the entire search session, even if a newly added seed paper cites or is cited by it unless the deleted item is recovered from the deleted paper list. It is effective to avoid repeatedly browsing identified irrelevant items.

Interactive Query Refinement. In traditional academic search systems, engagement of searchers in deciding computation of the relevance is limited. relevance can only be determined by keyword queries; positive and negative feedbacks are not utilized for improving results immediately. Therefore, PaperPoles provides user-driven methods for refining queries. The proposed interactions are designed to enable searchers to seamlessly utilize what they learn during the exploration to iteratively express their evolving information needs. The searchers can refine their queries by adding seed paper queries, adding negative queries, and tuning the weights of the queries.

Seed paper addition. A searcher adds a seed paper by dragging and dropping a paper node to the positive drop box or by clicking on the 'Thumbs Up' button in the list view. After a new seed paper is added, the searcher can decide when to load the citations by clicking the button of 'reference' and 'citation' (Figure 5). The cluster view would show the changes of search results. Loading citations is optional. A seed paper can also work as a positive query without citations loaded. The relevance scores of papers in the search results are recalculated. Papers being relevant to the newly added seed paper gains an increment of relevance score and move toward the left side.

Negative query addition. Searchers add a negative query by dragging and dropping a paper node to the negative drop box or by clicking on the 'Thumbs Down' button in the list view. Adding a negative query makes papers being relevant to the negative query move toward the right side.

Seed paper and negative query deletion. Searchers can also remove seed papers and negative queries by clicking on the close button (Figure 5) on the query icons. After removing a seed paper, its citations are to be removed, except items that are cited by or cite other seed papers. The effects of the removed queries on defining the relevance scores of papers are cleared.

Weight tuner. Searchers can adjust the weight of a query in relevance computation by tuning the slider (Figure 5) when hovering the query icon. When having multiple queries, the queries may play different roles in characterizing the information needs of the searchers. The weight tuner can help the searchers express their varying preferences for the queries.

Figure 5. Query tags with weight slider, delete button, and tooltip.

Experiment

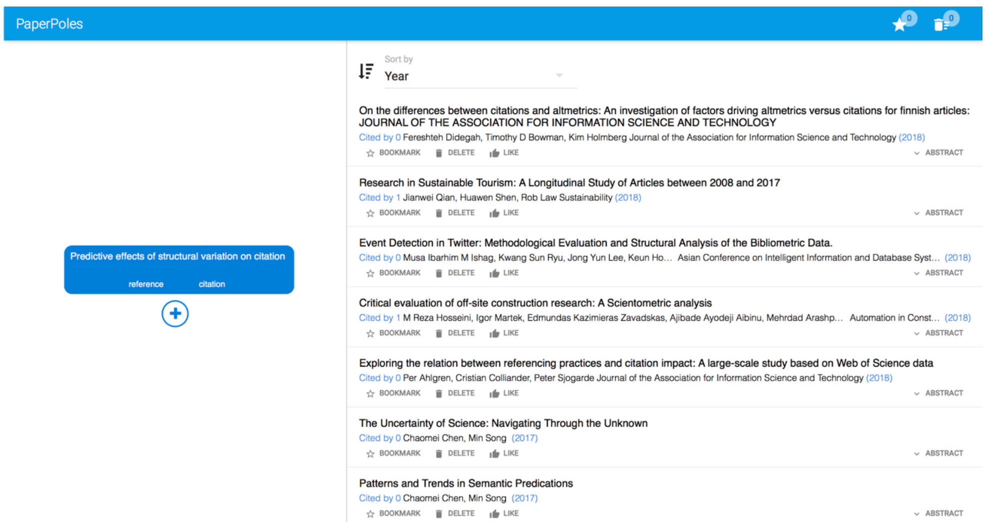

Figure 6. Baseline interface.

We conducted a user study to compare PaperPoles to a baseline system in two pre-defined tasks. The baseline (see Figure 6) was implemented to enable comparability with PaperPoles. The baseline does not feature the cluster view and the relevance ranking based on positive queries and negative queries.

We set up the experiment and designed measures to answer the following research questions:

1. Task performance: Can PaperPoles lead to better task performance?

2. Effectiveness of exploration: Can PaperPoles guide searchers to explore search result subsets meeting their information needs? In other words, can searchers identify a relevant paper by exploring fewer papers in PaperPoles?

3. Action difference: What are the differences of the usage actions between the systems and between the tasks?

Experiment Design

In the recruitment of participants, every candidate was given a prior background survey to collect their background information for task assignment and eligibility purposes.

We chose a 2×2 between-subjects experiment design with two tasks, T1 and T2, and two systems, Baseline (BL) and PaperPoles (PP). Each participant conducted the two tasks, but the tasks were conducted in different systems. Thus, a participant used each system and conducted each task only once, in order to avoid learning effects. We had four combinations of tasks and systems that are (1) BL+T1 and PP+T2; (2) BL+T2 and PP+T1; (3) PP+T1 and BL+T2; (4) PP+T2 and BL+T1. We varied the ordering to counterbalance the experiment conditions. We also used stratified randomization to divide participants into four groups by their background information (familiarity with the topics of tasks, research experience, and gender) to counterbalance their background effects. Each group was assigned by one of the four task combinations.

Prior to each task, the participants were asked to read written instructions explained the system and the task. Then the participants followed a tooltip-like step-by-step guide in the web-based interfaces of the baseline and PaperPoles. After the guide, they had 3-minute trial using the system via a predefined query in the system. After the trial, the system would be reset for the actual experiment and data collection.

A post-task survey was given to the participants after each task to collect data about their subjective experience.

Data and Participants

We used Microsoft Academic Graph as the data source that offers over 176 million scientific publications and has a broad coverage of various disciplines. We access the dataset by Microsoft Cognitive Services Academic Knowledge API, through which we can request the following information of articles: title, abstract, authors, venue, keywords, references, publication year, and citations.

57 participant candidates responded to our prior background survey. By the survey, we selected participants from candidates by the following criteria. The information about the skills and knowledge of the candidates were reported by their self-assessment.

1. They had experience of using academic search systems;
2. They had good English knowledge for reading English papers;
3. They either had a graduate-level degree or were pursuing a graduate-level degree in Library and Information Science;
4. They were "slightly familiar" (2 out of 5 points) at least to the task topics.

The participants selected by using these criteria should be able to search English papers about the two task topics by using academic search systems.

Eventually, we recruited 28 participants, including 2 full professors, 4 associate professors, 6 assistant professors, 12 doctoral students, and 4 master students. 16 (58%) of them were females. The participants received financial compensation after the experiment.

We assigned 7 participants for each combination of tasks by their task topic familiarity, research experience, and genders. Thus, we have 14 participants with a similar background distribution for each task on each system. The self-reported familiarity to task topics (5-point scale) of participants is summarized in Table 1. The participants for both tasks had no significant difference between the systems ($p > 0.05$).

Table 1
Topic familiarity of participants.

	Baseline		PaperPoles	
	M (SE)	*N*	M (SE)	*N*
T1	3.93 (0.16)	14	3.82 (0.18)	14
T2	3.24 (0.31)	14	3.10 (0.40)	14

Search Tasks

The work scenario was: "You are searching related work of your research. You have identified some relevant papers and you are planning to identify other papers based on the references and citations of these seed paper".

The topics of tasks should be understandable for researchers in the fields of Library and Information Science. Seed papers should have enough citations to be explored. The tasks should have different levels of complexity so that we can evaluate the performance improved by PaperPoles on tasks with different complexity. We chose "factors affecting citations" (T1) and "science mapping" (T2) as the task topics. Both of them are well-studied research topics in Library and Information Science. Participants were asked to complete each task in 20 minutes, but they were allowed to spend more than 20 minutes.

- **T1**. Participants were given one seed paper (C. Chen, 2012) that studied a factor affecting citation counts of research articles. The participants were asked to identify 10 papers that also studied factors affecting citation.
- **T2**. Participants were given two seed papers (Cobo et al., 2012; Wang, Cheng, & Lu, 2014) that described two visualization tools for science mapping. The participants were asked to identify 10 papers that studied other visualization methods and tools, but not including case studies conducted by using science mapping tools nor visualization methods and tools that were not designed specifically for science mapping.

The levels of difficulty of the two tasks are different. T2 is difficult than T1 because science mapping is a highly interdisciplinary research area related to information visualization, scientometrics, network science, and various computational areas. The interdisciplinarity leads to the diverse citations of the seed papers. Additionally, T2 has more specific criteria than T1. The results of task performance (see Results section) also show the differences between two tasks regarding the level of difficulty.

Measures

Task performance. We measured two aspects of task performance: efficiency and effectiveness. They were measured by the time participants spent to complete the tasks and the accuracy of results searched by participants respectively.

- **Completion Time**. We recorded duration from the beginning of a task to the completion to measure completion time.

- **Accuracy**. We measured effectiveness as the precision of the papers identified by participants. All papers identified for each task by any of the participants in any of systems were pooled. Two experts then assessed the actual relevance of each paper to the topic of the task on a binary scale: 0 (not-relevant) or 1 (relevant). The Cohen Kappa test indicated substantial agreement between two experts, Kappa = 0.81, $p < 0.001$. The disagreement between two experts was resolved by the third expert. The experts were listed as authors of this paper.

Effectiveness of exploration. The cluster view is designed to help searchers identify a set of papers of interest and relevance for further exploration. If the view is effective for this goal, participants using PaperPoles would explore fewer papers to identify a relevant paper on average than using the baseline, i.e., the effectiveness of their exploration is improved by PaperPoles.

The explored papers include the paper items rendered and displayed for more than two seconds in the list view as well as the paper nodes that mouse pointer was over for more than two seconds in the cluster view. The exploration effectiveness was measured as the number of identified relevant papers over the number of explored articles.

The differences in usage actions. We recorded usage actions for comparing the usage actions for two systems and two tasks. Both exploration and control actions were recorded.

Results

Task Performance

The results of the experiment regarding task performance are summarized in Table 2 and details were illustrated in Figure 7 and Figure 8 with respect to the measures of completion time and accuracy.

Completion time. The significant difference of task completion time between the systems was found in T2. The results show that participants spent substantially less time completing T2 when using PaperPoles than when using the baseline. The mean task duration for PaperPoles was 12.4 minutes, while the mean task duration for the baseline system was 17.5 minutes. The differences between the systems were found statistically significant by using paired sample t-test ($p < 0.05$).

However, we also found no significant difference between the systems in T1. The reason may be the different difficulty levels of the tasks. The time spent to complete T1 is much less than the time for T2 in the systems, and the accuracy of the papers identified by the participants in T2 was much lower than in T1 in the systems. T2 is more difficult because of its interdisciplinary nature, as mentioned earlier. The results confirmed this. Thus, the results suggest that PaperPoles can improve the efficiency of completing complex search tasks, but in relatively simple tasks, the improvement may not be significant.

To unfold the whole process of completing the tasks, the mean accumulated time spent by the participants over each paper (from first to tenth papers) in the systems is displayed in Figure 7. In T1, the increasing trends of time spent by the participants on the systems are very similar and the

Table 2
Task performance summary.

		Baseline (BL)	PaperPoles (PP)	BL vs. PP
Mesaurement	Task	M (SE)	M (SE)	p
Completion Time (seconds)	T1	487 (50)	475 (70)	0.834
	T2	1,047 (167)	**742** (112)	**0.035**
Accuracy	T1	0.80 (0.04)	**0.91** (0.05)	**0.048**
	T2	0.28 (0.05)	**0.41** (0.03)	**0.022**

difference of the time spent in the systems are not significant all the time. However, in T2, we see a different increasing trend in the systems. The time spent in the baseline system has a steeper increasing trend after the sixth papers. The difference of the time spent for T2 between the systems is significant at the first paper and after the sixth paper. PaperPoles can prevent the search process of a complex task becoming increasingly slow.

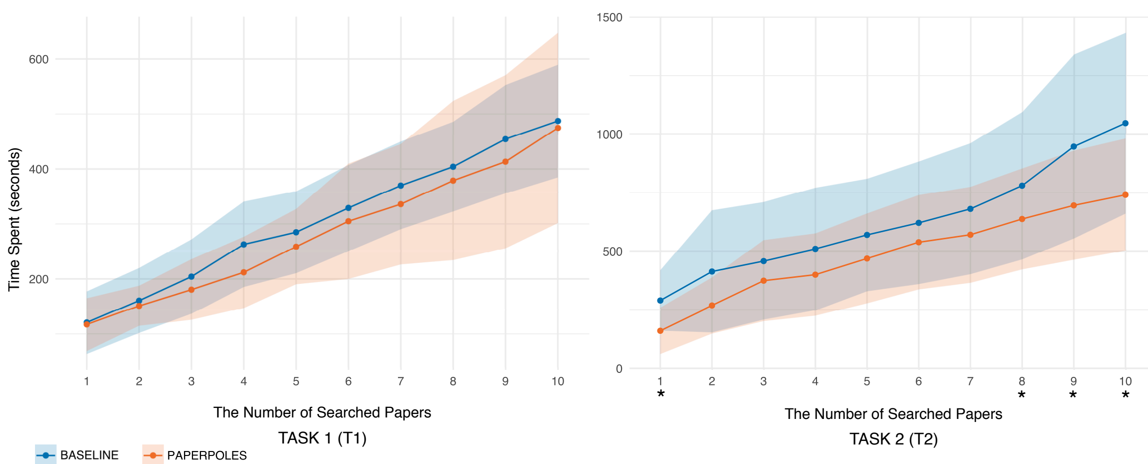

Figure 7. The time spent for searching the first to tenth papers. The data of time spent in completing the tasks on the systems with 95% confidence intervals are displayed. (Statistical significance level: (*) $<$ 0.05)

Accuracy. Significant differences in the measured accuracy between the systems were found on both tasks. In T1, mean accuracy on the baseline system was 0.8, but mean accuracy on PaperPoles was 0.91. In T2, the difference between the systems was greater (0.28 vs. 0.41). PaperPoles was more effective for both of the simple task and complex task than the baseline.

Exploration Effectiveness

Table 3 shows the exploration effectiveness and the data that were used for computing the effectiveness. In T1, the number of viewed papers differed significantly between the systems, but the exploration effectiveness had no significant difference between them. In T2, the number of viewed papers was not significantly different between the systems, but the number of accurate searched

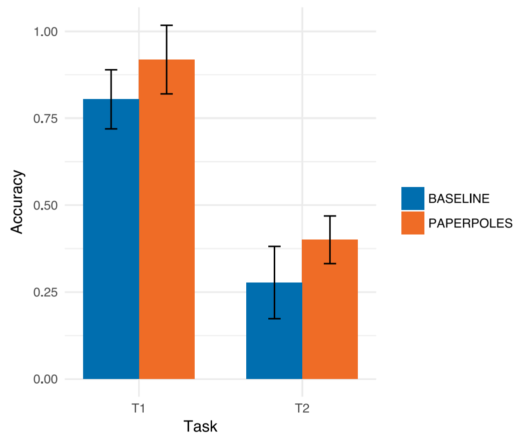

Figure 8. Accuracy displayed for both systems with 95% confidence intervals.

papers and the effectiveness of exploration were significantly different between them. Thus, PaperPoles could significantly improve the effectiveness for the complex task but PaperPoles might not improve it for the simple task.

Table 3
Exploration effectiveness.

		Baseline (BL)	PaperPoles (PP)	BL vs. PP
Task	Dimensions	M (SE)	M (SE)	p
T1	Viewed Articles in List	55.7 (6.6)	70.1 (15)	0.113
	Viewed Papers in Clusters	-	9.5 (4.9)	-
	Total Viewed Papers	55.7 (6.6)	79.6 (13.6)	**0.011**
	Retrieved Papers	14.9 (2.0)	12.9 (2.2)	0.474
	Accurate Searched Papers	12.1 (1.9)	11.7 (1.6)	0.896
	Exploration Effectiveness	0.24 (0.02)	0.21 (0.06)	0.468
T2	Viewed Papers in List	63.1 (9.2)	28.3 (4.1)	**0.004**
	Viewed Papers in Clusters	-	31.0 (7.0)	-
	Total Viewed Papers	63.1 (9.2)	59.3 (13.1)	0.869
	Retrieved Papers	14.1 (1.9)	15.2 (2.0)	0.667
	Accurate Searched Papers	3.7 (0.8)	6.3 (1.1)	**0.045**
	Quality of Viewed Information	0.07 (0.01)	0.11 (0.02)	**0.038**

User Actions

Table 4 shows the system usage of two systems. In both tasks, the participants in PaperPoles added seed papers more than in the baseline significantly, but there was no significant difference in the action of loading citations between the systems. The participants on PaperPoles tended to remove seed papers more, but the difference was not significant.

In T2, the participants tended to use various actions to refine their search result set. For example, the participants added seed papers more in both the systems; they also loaded citations, added negative seed papers, tuned the query weights, and selected nodes more frequently for T2 in PaperPoles. This finding is not surprising because the cumbersome search process in complex tasks encouraged the participants to use more types of actions.

Table 4
Usage action summary.

		Baseline (BL)		PaperPoles (PP)		BL vs. PP
Task	Action	M (SE)	Searcher Count	M (SE)	Searcher Count	p
T1	Add Seed Paper	5.7 (1.1)	14	9.4 (1.4)	14	**0.018**
	Add Negative Paper	-	-	3.1 (1.3)	6	-
	Remove Seed Paper	2.3 (1.3)	4	6.0 (2.7)	5	0.351
	Remove Negative Paper	-	-	8 (0)	1	-
	Load Citations	6.4 (0.8)	14	3.9 (0.9)	14	0.255
	Tune Weight	-	-	8.0 (0)	2	-
	Select Nodes	-	-	1.2 (0.8)	9	-
T2	Add Seed Paper	7.8 (1.1)	14	12.1 (1.6)	14	**0.036**
	Add Negative Paper	-	-	6 (1.9)	8	-
	Remove Seed Paper	2.8 (0.8)	5	4.0 (1.4)	5	0.170
	Remove Negative Paper	-	-	1.0 (0)	2	-
	Load Citations	6.7 (0.9)	14	5.4 (0.8)	14	0.229
	Tune Weight	-	-	5.0 (1.6)	9	-
	Select Nodes	-	-	2.8 (1.1)	13	-

Subjective Feedback Analysis

To compare subjective feedback over the systems and the tasks, we analyzed the responses of the post-task questionnaire. The analyzed aspects in the questionnaire were selected from the questionnaire used by Tsai and Brusilovsky (2018) and ResQue questionnaire (Pu, Chen, & Hu, 2011). Figure 9 shows the results of the analysis. Due to the different nature of the tasks, we analyzed the results of the two tasks separately. In T1, the baseline system received a significantly higher rating on Easy to Use (Q6) and Enjoyable (Q7). PaperPoles did not receive any significantly higher rating. PaperPole only received a slightly higher rating on Usefulness (Q2), but it was not significant. The results in T2 were different. In T2, the PaperPoles received a higher rating on five aspects: Usefulness (Q2), Satisfaction (Q4), Intention to Reuse (Q8), Enjoyable (Q9), and No Real Benefit (Q10).

It is interesting to see the opposite ratings for Easy to Use and Enjoyable between the tasks. The feedback shows that the baseline system was significantly easier to use than PaperPoles in T1, but PaperPoles was rated to be easier to use (though not significantly so). Besides, both systems received significantly opposite ratings for Enjoyable when the systems were used in the different tasks. The participants might feel more enjoyable and even think easier to use when they found the features of the system they used were useful for completing their task. This finding is consistent with the findings found in the analysis of completion time and usage actions.

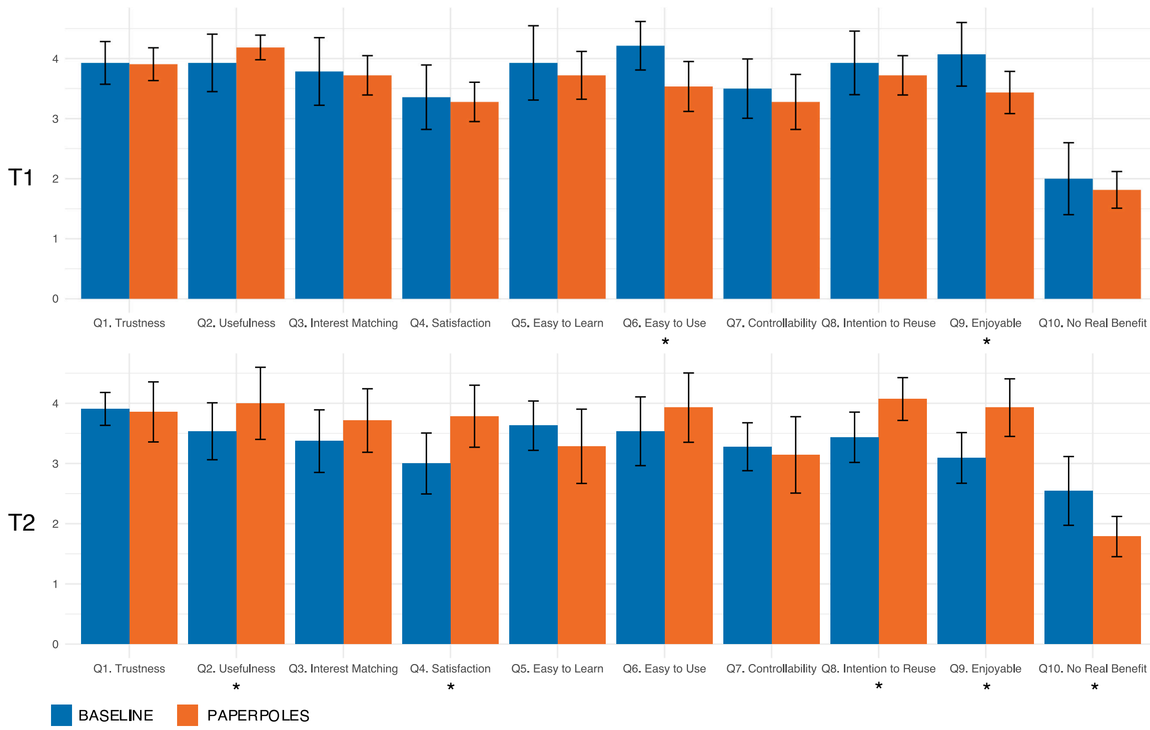

Figure 9. Feedback analysis results with confidence interval. (Statistical significance level: (*) < 0.05)

Discussion

We built a context-aware environment for publication exploration through an interface with multiple coordinated views and a workflow fitting evolving information needs of users. We also conducted an experiment to investigate the task performance, exploration effectiveness, and action patterns of the participants for completing two tasks on the PaperPoles and the baseline system. We discuss the experiment results, limitations of this paper, and implications in this section.

1. Can PaperPoles lead to better task performance?

   The results of the experiment show PaperPoles can significantly improve effectiveness in both simple task (T1) and complex task (T2), without compromising efficiency. In the complex task, the efficiency can be improved by PaperPoles compared with the baseline system, but, in the simple task, no statistically significant improvement was found in PaperPoles.

   Participants could use PaperPoles to complete both tasks with higher accuracy. The interface of PaperPoles allowed participants to describe their information needs and refine their queries intuitively so that participants can identify papers fitting their information needs regardless of the task complexity. Therefore, PaperPoles can improve the effectiveness for tasks with different complexity.

The experiment also shows that the participants could not improve their efficiency performance in simple tasks. Our visualization allowed searchers to convey their complex information needs intuitively and have a good understanding of semantic aspects and the distribution of relevant items. Thus, the participants were able to quickly narrow down to a result subset fitting their complex criteria. However, the efficiency of conducting a simple task may not be improved by the visualization, because the narrowing-down process also costs time and may not be worthwhile in simple tasks.

2. Can PaperPoles guide searchers to explore search result subsets meeting their information needs?

   In exploratory search, searchers usually issue a tentative and imprecise query to navigate proximal to relevant documents. PaperPoles can guide searchers to explore the search result by navigating to a result subset that potentially has more relevant papers so that user may explore fewer irrelevant papers, i.e., PaperPoles may improve the exploration effectiveness. However, the effectiveness can be improved in PaperPoles only for a complex task. A possible reason for no improvement of the effectiveness in the simple task is that participants were less likely to use the visualization and interaction features on PaperPoles when conducting a simple task. The usage pattern in the simple task was supported by the findings in the analysis of usage actions.

3. What are the differences in the usage actions between the systems and between the tasks?

   The participants tended to add and remove seed papers more on PaperPoles than the baseline system, which may be due to the controllability and transparency of PaperPoles. In PaperPoles, participants could see the changes brought by adding or removing seed paper and can also refine the queries, so they might have the confidence to keep the search in control after taking the actions.

   We also found that participants were more likely to take various actions in the complex task than in the simple task. The complexity of a task may motivate searchers to take possible actions that may improve the search efficiency.

The experiment results can provide an explanation of when searchers need visualization interfaces, how/why they can benefit from visualization interfaces, and what the benefits are. In the light of the results, we can provide the implications for designing interfaces and interactions in an exploratory search system and understanding exploratory behavior in academic search.

The design of PaperPoles may have implications for designing interfaces and interactions in an exploratory search system. The cluster view in PaperPoles shows an aspect overview and relevance patterns of search results, which can help searchers identify a subset meeting their information needs well to explore. Searchers may avoid exploring subsets with too many irrelevant results. However, the information needs of searchers may be evolving and uncertain. For example, when a research start an exploratory search on an unfamiliar research area, the researcher may not have specific information needs. The interactions in PaperPoles allow users to express their information needs based on what the new information they consumed in the exploration. Based on our study, enabling searchers to intuitively and iteratively express their information needs in an exploration process can improve their performance on a complex search task.

The experiment results shed light on understanding exploratory behavior in academic search. The complexity level of a task plays a critical role in the process of an exploratory search. Although visualization interfaces are useful for various tasks, using them requires time for interactions and understanding the visualization. When searchers face the tradeoff between using a conventional interface and a complex visualization interface, the complexity level of tasks may be an important factor affecting their preferences. Besides, our results also indicate the extent and types of benefits of using visualization are different for tasks with different complexity.

Not all the benefits brought by an interface can be perceived by searchers. The efficiency improvement was easily perceived by searchers, but they may not perceive the effectiveness improvement. The subjective feedback shows that the participants in the simple task (T1) scored PaperPoles lower than the baseline system on many evaluation aspects. However, the accuracy performance of T1 on PaperPoles was significantly higher than in the baseline and the completion time was not more than on the baseline system.

While the idea of enabling searchers to express their evolving information needs in a context-aware environment is promising, this paper has limitations. Some possibly important information is not displayed or is displayed in a less explicit form. For example, the relevance of a searched item is determined by multiple visually encoded queries, but the visualization does not explicitly reveal the origins of the relevance of an item and lacks an interaction to allow searchers to reason the details; The relations among papers and relations among clusters are not displayed. The animation and visual design may be improved to help searchers understand the changes brought by a newly added query. The text analysis and network analysis techniques may not perform well on a small search result set. We used a list-based interface as the baseline to conduct a comparative experiment, but interfaces enhanced by visualization techniques (Mackinlay et al., 1995; Ponsard et al., 2016) which may also support similar tasks were not compared.

Conclusion

This paper presents a visual analytics system, PaperPoles, to help searchers interactively search for papers of interest based on citation links of papers that are known to be relevant. In PaperPoles, searchers can intuitively form positive and negative queries and act on search results immediately by adjusting an iterative query expansion process. PaperPoles also enables searchers to explore a set of search results by topic and by relevance.

We also conducted an experiment that evaluated the usability of the system and developed an understanding of how searchers benefit from using the system. The results of the experiment suggest that the system can improve task effectiveness significantly over simple and complex tasks. However, it can improve the task efficiency and the exploration effectiveness only over the complex task. The analysis of usage actions and subjective feedback suggest that searchers are more likely to take advantage of the cluster visualization and query refinement interaction to improve their performance in the complex task.

In the future, we plan to improve system performance by improving interaction design and underlying techniques including the clustering method and text analytic techniques. We also plan to add animations and visual elements to increase the transparency of the system. Another exciting avenue for future work is to apply the design on generic academic exploratory search and generic document exploratory search based on the design rationale of the proposed system and findings from the experiment. We believe the visualization and interaction design can also benefit other types of search tasks. In the future, we will release a stable version online for free use and collect user log data for a longitudinal user study.

Acknowledgment

Jiangen He, Qing Ping, and Chaomei Chen wish to thank the support of the National Science Foundation (Award Number: 1633286). Wen Lou wishes to thank the support of the National Social Science Foundation of China (Award Number: 17CTQ025)

The authors thank Andrea Forte, Swathi Jagannath, Ali Jazayeri, Adam Johs, Kai Li, Meaghan Lutts, and Wei Quan at Drexel University for their suggestions and comments on the study and writing, and also thank Changyang Feng, Xin Li, and Ye Chen at Wuhan University for their help in the participant recruitment. We also appreciate the constructive comments from the anonymous reviewers.

References

Ahn, J., & Brusilovsky, P. (2013). Adaptive visualization for exploratory information retrieval. *Information Processing & Management*, *49*(5), 1139–1164.

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

Athukorala, K., Hoggan, E., Lehtiö, A., Ruotsalo, T., & Jacucci, G. (2013). Information-seeking behaviors of computer scientists: Challenges for electronic literature search tools. *Proceedings of the Association for Information Science and Technology*, *50*(1), 1–11.

Athukorala, K., Micallef, L., An, C., Reijonen, A., Peltonen, J., Ruotsalo, T., & Jacucci, G. (2017). Visualizing activity traces to support collaborative literature searching. In *Proceedings of the 10th International Symposium on Visual Information Communication and Interaction* (pp. 45–52). New York, NY, USA: ACM.

Baldonado, M. Q. W., & Winograd, T. (1997). Sensemaker: An information-exploration interface supporting the contextual evolution of a user's interests. In *Proceedings of the acm sigchi conference on human factors in computing systems* (pp. 11–18). New York, NY, USA: ACM. doi: 10.1145/258549.258563

Barrett, A. (2005). The information-seeking habits of graduate student researchers in the humanities1. *The Journal of Academic Librarianship*, *31*(4), 324–331.

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

Beck, F., Koch, S., & Weiskopf, D. (2016). Visual analysis and dissemination of scientific literature collections with survis. *IEEE Transactions on Visualization and Computer Graphics*, *22*(1), 180–189.

Belkin, N. J., et al. (1993). Interaction with texts: Information retrieval as information seeking behavior. *Information retrieval*, *93*, 55–66.

Berger, M., McDonough, K., & Seversky, L. M. (2017). cite2vec: Citation-driven document exploration via word embeddings. *IEEE Transactions on Visualization and Computer Graphics*, *23*(1), 691-700.

Blondel, V. D., Guillaume, J., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), 155–168.

Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, *61*(12), 2389–2404.

Chau, D. H., Kittur, A., Hong, J. I., & Faloutsos, C. (2011). Apolo: Making sense of large network data by combining rich user interaction and machine learning. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 167–176). New York, NY, USA: ACM. doi: 10.1145/1978942.1978967

Chen, C. (2006). Citespace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, *57*(3), 359–377.

Chen, C. (2012). Predictive effects of structural variation on citation counts. *Journal of the American Society for Information Science and Technology*, *63*(3), 431-449.

Chen, T., & Gross, M. (2017). Usability modeling of academic search user interface. In *Design, user experience, and usability: Understanding users and contexts* (pp. 16–30). Springer International Publishing.

Choo, J., Kim, H., Clarkson, E., Liu, Z., Lee, C., Li, F., . . . Park, H. (2018, January). Visirr: A visual analytics system for information retrieval and recommendation for large-scale document data. *ACM Trans. Knowl. Discov. Data*, *12*(1), 8:1–8:20. doi: 10.1145/3070616

Clarkson, E. C., Desai, K., & Foley, J. D. (2009). ResultMaps: Visualization for search interfaces. *IEEE Transactions on Visualization and Computer Graphics*, *15*(6), 1057–1064.

Cobo, M., López-Herrera, A., Herrera-Viedma, E., & Herrera, F. (2012). SciMAT: A new science mapping analysis software tool. *Journal of the American Society for Information Science and Technology*, *63*(8), 1609–1630.

di Sciascio, C., Sabol, V., & Veas, E. E. (2016). Rank as you go: User-driven exploration of search results. In *Proceedings of the 21st international conference on intelligent user interfaces* (pp. 118–129). New York, NY, USA: ACM.

Dork, M., Riche, N. H., Ramos, G., & Dumais, S. (2012, Dec). Pivotpaths: Strolling through faceted information spaces. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2709-2718. doi: 10.1109/TVCG.2012.252

Du, J. T., & Evans, N. (2011). Academic users' information searching on research topics: Characteristics of research tasks and search strategies. *The Journal of Academic Librarianship*, *37*(4), 299–306.

Dunne, C., Shneiderman, B., Gove, R., Klavans, J., & Dorr, B. (2012). Rapid understanding of scientific paper collections: Integrating statistics, text analytics, and visualization. *Journal of the American Society for Information Science and Technology*, *63*(12), 2351–2369.

Ellis, D. (1989). A behavioural model for information retrieval system design. *Journal of information science*, *15*(4-5), 237–247.

Ellis, D., Cox, D., & Hall, K. (1993). A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of documentation*, *49*(4), 356–369.

Federico, P., Heimerl, F., Koch, S., & Miksch, S. (2017). A survey on visual approaches for analyzing scientific literature and patents. *IEEE Transactions on Visualization and Computer Graphics*, *23*(9), 2179-2198.

Felix, C., Pandey, A. V., & Bertini, E. (2017). Texttile: An interactive visualization tool for seamless exploratory analysis of structured data and unstructured text. *IEEE Transactions on Visualization and Computer Graphics*, *23*(1), 161-170.

Fox, E. A., & Shaw, J. A. (1994). Combination of multiple searches. *NIST special publication SP*, 243–252.

Görg, C., Liu, Z., Kihm, J., Choo, J., Park, H., & Stasko, J. (2013). Combining computational analyses and interactive visualization for document exploration and sensemaking in jigsaw. *IEEE Transactions on Visualization and Computer Graphics*, *19*(10), 1646–1663.

Hearst, M. A. (1995). Tilebars: Visualization of term distribution information in full text information access. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 59–66).

Heimerl, F., Han, Q., Koch, S., & Ertl, T. (2016). Citerivers: Visual analytics of citation atterns. *IEEE Transactions on Visualization and Computer Graphics*, *22*(1), 190–199.

Hoeber, O., & Yang, X. D. (2006). The visual exploration of web search results using hotmap. In *Proceedings* *of the Information Visualization (IV'06)* (pp. 157–165).

Iwata, M., Sakai, T., Yamamoto, T., Chen, Y., Liu, Y., Wen, J.-R., & Nishio, S. (2012). Aspectiles: Tile-based visualization of diversified web search results. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 85–94).

Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, *28*(1), 11-21.

Kairam, S., Riche, N. H., Drucker, S., Fernandez, R., & Heer, J. (2015). Refinery: Visual exploration of large, heterogeneous networks through associative browsing. *Computer Graphics Forum*, *34*(3), 301-310.

Kessler, M. (1963). Bibliographic coupling between scientific papers. *American documentation*, *14*(1), 10–25.

Khazaei, T., & Hoeber, O. (2017). Supporting academic search tasks through citation visualization and exploration. *International Journal on Digital Libraries*, *18*(1), 59–72.

Kim, J. (2009). Describing and predicting information-seeking behavior on the web. *Journal of the Association for Information Science and Technology*, *60*(4), 679–693.

Klouche, K., Ruotsalo, T., Micallef, L., Andolina, S., & Jacucci, G. (2017). Visual re-ranking for multi-aspect information retrieval. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval* (pp. 57–66). ACM.

Lee, B., Czerwinski, M., Robertson, G., & Bederson, B. B. (2005). Understanding research trends in conferences using paperlens. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (pp. 1969–1972). New York, NY, USA: ACM.

Li, X., Schijvenaars, B. J., & de Rijke, M. (2017). Investigating queries and search failures in academic search. *Information Processing & Management*, *53*(3), 666 - 683.

Mackinlay, J. D., Rao, R., & Card, S. K. (1995). An organic user interface for searching citation links. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '95* (pp. 67–73).

Marchionini, G. (2006). EXPLORATORY SEARCH: From finding to understanding. *Communications of the ACM*, *49*(4), 41–46.

Meho, L. I., & Tibbo, H. R. (2003). Modeling the information-seeking behavior of social scientists: Ellis's study revisited. *Journal of the Association for Information Science and Technology*, *54*(6), 570–587.

Montague, M., & Aslam, J. A. (2001). Relevance score normalization for metasearch. In *Proceedings of the 10th International Conference on Information and Knowledge Management* (pp. 427–433).

Nguyen, T. N., & Zhang, J. (2006). A novel visualization model for web search results. *IEEE Transactions on Visualization and Computer Graphics*, *12*(5), 981–988.

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

Palla, G., Barabási, A., & Vicsek, T. (2007). Quantifying social group evolution. *Nature*, *446*(7136), 664–667.

Peltonen, J., Belorustceva, K., & Ruotsalo, T. (2017). Topic-relevance map: Visualization for improving search result comprehension. In *Proceedings of the 22nd international conference on intelligent user interfaces* (pp. 611–622). New York, NY, USA: ACM.

Ponsard, A., Escalona, F., & Munzner, T. (2016). Paperquest: A visualization tool to support literature review. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2264–2271). New York, NY, USA: ACM.

Pu, P., Chen, L., & Hu, R. (2011). A user-centric evaluation framework for recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems* (pp. 157–164). New York, NY, USA: ACM.

Rhee, H. L. (2012). Modelling historians' information-seeking behaviour with an interdisciplinary and comparative approach. *Information Research*, *17*(4), paper544.

Ruotsalo, T., Peltonen, J., Eugster, M., Glowacka, D., Konyushkova, K., Athukorala, K., ... Kaski, S. (2013). Directing exploratory search with interactive intent modeling. In *Proceedings of the 22Nd ACM In-*

ternational Conference on Information & Knowledge Management (pp. 1759–1764). New York, NY, USA: ACM.

Savolainen, R. (2017a). Berrypicking and information foraging: Comparison of two theoretical frameworks for studying exploratory search. *Journal of Information Science*, *43*(3), 1–14.

Savolainen, R. (2017b). Contributions to conceptual growth: The elaboration of ellis's model for information-seeking behavior. *Journal of the Association for Information Science and Technology*, *68*(3), 594–608.

Smith, G., Czerwinski, M., Meyers, B., Robbins, D., Robertson, G., & Tan, D. S. (2006). FacetMap: A scalable search and browse visualization. *IEEE Transactions on Visualization and Computer Graphics*, *12*(5), 797–804.

Teevan, J., Alvarado, C., Ackerman, M. S., & Karger, D. R. (2004). The perfect search engine is not enough: A study of orienteering behavior in directed search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 415–422). New York, NY, USA: ACM.

Toutanova, K., & Manning, C. D. (2000). Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In *Proceedings of the 2000 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora: Held in Conjunction with the 38th Annual Meeting of the Association for Computational Linguistics* (pp. 63–70).

Tsai, C. H., & Brusilovsky, P. (2018). Beyond the ranked list: User-driven exploration and diversification of social recommendation. In *23rd international conference on intelligent user interfaces* (pp. 239–250). New York, NY, USA: ACM.

Umemoto, K., Yamamoto, T., & Tanaka, K. (2016). Scentbar: A query suggestion interface visualizing the amount of missed relevant information for intrinsically diverse search. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 405–414).

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538.

Wang, X., Cheng, Q., & Lu, W. (2014). Analyzing evolution of research topics with neviewer: a new method based on dynamic co-word networks. *Scientometrics*, *101*(2), 1253–1271.

White, R. W., & Roth, R. A. (2009). *Exploratory search: Beyond the query-response paradigm*. San Rafael, CA: Morgan & Claypool Publishers.

Zhao, J., Collins, C., Chevalier, F., & Balakrishnan, R. (2013, Dec). Interactive exploration of implicit and explicit relations in faceted datasets. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2080-2089. doi: 10.1109/TVCG.2013.167

Zhu, Y., & Yan, E. (2016). Searching bibliographic data using graphs: A visual graph query interface. *Journal of Informetrics*, *10*(4), 1092 - 1107.