Exploratory Search: A Critical Analysis of the Theoretical Foundations, System Features, and Research Trends

Tingting Jiang

Abstract Humans are explorers by nature. Almost all searches are exploratory to a certain extent. As a result of the subdivision of the information seeking domain, exploratory search has become a new research focus arousing extensive attention. This chapter introduces the concept of exploratory search and illustrates its basic theoretical foundations, clarifying its complex meaning from the aspects of the problem context and the search process. Four different methods of classifying search results are identified based on a survey of existing exploratory search systems, including hierarchical classification, faceted classification, dynamic clustering, and social classification. Their inherent characteristics and practical applications are reviewed in detail, and the visualization support for presenting the classified search results is explored in addition. The development trends of the exploratory search field are predicted according to the social nature of information seeking.

Keywords Exploratory search · Systems · Users · Classification · Visualization · Social

Introduction

Online information seeking is an indispensable part of our daily lives and work. Explicit information needs, e.g. the needs for weather, flight, and stock information, can be quickly satisfied by powerful Web search engines. This reflects the look-up model that focuses on matching user queries with document surrogates (Bates 1989). Specific queries will lead to accurate search results, and one does not need to make any evaluation or comparison (Marchionini 2006).

Tingting Jiang is currently an associate professor at the School of Information Management, Wuhan University. She received her PhD in Library and Information Science from the University of Pittsburgh. Her current research interests include information seeking behavior, information architecture, information visualization, and Web 2.0. Tingting Jiang can be contacted at: tij@whu.edu.cn.

T. Jiang (✉)
School of Information Management, Wuhan University, Wuhan, China
e-mail: tij@whu.edu.cn

   Nevertheless, the lookup model is not applicable to many real-world scenarios. Scientific researchers may want to dig into a new research topic; budget travelers may want to make an affordable travel plan; youngsters may want to learn the secrets of career success; and so on and so forth. The information needs involved in these problems cannot be directly translated into appropriate queries, because people are not familiar with the knowledge domain that is related to the search, they do not know the means of achieving their goals, or the goals are not clear in themselves (Nolan 2008).

   In tackling the above problems, as a matter of fact, people have to define their search goals in the first place. The information they obtain at the beginning of the search process may be of poor relevance. However the more information they absorb, the more thoroughly they understand the problem. In this way people get to distinguish between what they already know and what they should know. The gap in between is the information need. With the need taking shape gradually, people will be more and more able to formulate queries and identify relevant items. At this moment, the power of the search system in automatic matching starts to play its role truly. Whether people can find satisfying solutions to the original problem is further dependent upon their skills of extracting valuable information from search results. Here we see the user-dominated non-linear search, known as "exploratory search".

   Exploratory search is a special type of information seeking. The 2005 Exploratory Search Interface Workshop was the first milestone in the history of this sub discipline (White et al. 2005). It was followed by a series of influential events, including the 2006 ACM SIGIR Workshop on Evaluating Exploratory Search Systems, the 2007 ACM SIGCHI Workshop on Exploratory Search and HCI, and the 2008 NSF Invitational Workshop on Information Seeking Support Systems. Moreover several academic journals, such as the *Communications of the ACM, the International Journal of Information Processing and Management,* and *Computer*, have published special issues on exploratory search.

Related Work

Classical Theories Related to Exploratory Search

Many researchers from the areas of information retrieval, human-computer interaction, information organization, and information behavior have devoted their attention to exploratory search. Indeed, exploratory search studies can seek theoretical roots in these areas. Below is a brief review of frequently cited related theories from two aspects, i.e. users' internal cognition and external behavior.

Interactive Information Retrieval and Cognitive Information Retrieval

Interactive information retrieval changes the system-centered tradition adopted by early information retrieval research and concentrates more on the user's input and

control in the search process. It is closely related to cognitive information retrieval because the main purpose of interaction is to influence the user's cognitive state to make him/her more effective in information searching (Saracevic 1996).

As Ingwersen (1996) stated, all the interactive activities in information retrieval could arouse cognition processes. He created the polyrepresentations of both the information space of information retrieval systems and the cognitive space of users. While the former consists of the system setting and information objects, the latter includes four elements, i.e. work-task/interest domain, current cognitive state, problem space, and information need, which follow the bottom-up order of causality.

Similarly, Saracevic's (1997) stratified model also considers two sides: human and computer. In this model interaction is understood as a sequence of processes occurring at several levels, such as the cognitive, affective, and situational levels on the human side and the engineering, processing, and content levels on the computer side.

Guided by the hypothesis of anomalous states of knowledge (ASK), Belkin (1996) established the episode model in which an information seeking episode was defined as a series of interaction between the user and the information. The type of interaction at a certain time point is determined by the user's goals, intentions, situations, and the interaction is supported by such processes as representation, comparison, presentation, navigation, and visualization, etc.

The interactive feedback model by Spink (1997) resulted from an empirical study exploring how interaction occurred during mediated online searching. The search process may consist of multiple cycles, and multiple interactive feedback loops may be seen in each cycle. The interactive feedback covers the users' judgment regarding content relevance, term relevance, and magnitude as well as their review of tactics and terminologies.

Evolving Search and Information Foraging

Bates (1989) put forward two important arguments in her evolving search theory. First, users' query will keep changing in most real-world searches. Such changes may be not limited to term modifications. As the new information encountered in the search brings in new ideas, users' information needs will evolve. Second, an information need is not met by a single set of best results. Instead, the user will collect some useful information at each stage of the ever-modifying search, and the search goal is achieved by combining all these fragments. So to speak, evolving search follows the "berrypiking" pattern.

The theory of information foraging is more concerned with the evolution of search activities. There is an analogy between humans looking for information and animals looking for food in the nature. The best foragers are able to maximize the rate of valuable information acquired per unit cost. According to Pirolli and Card (1999), the task environment of information foraging presents a "patch" structure. Information is located in patches, and foragers assess the value of a patch in virtue of information scent, the perception of the patch gained from proximal cues.

In order to improve their efficiency in information foraging, people may try to lower the average costs of moving between the patches or increase the benefits of information acquisition in the current patch.

Important Efforts to Define Exploratory Search

Evolving search and information foraging emphasize the influences of environmental changes on users' search directions, whereas interactive information retrieval and cognitive information retrieval believe that users' subjective characteristics and the interaction objects (i.e. system or information) can affect each other given a specific search goal. These theories all play their roles in shaping the understanding of exploratory search by considering users' physical and mental functions in search. More recently, White and Roth (2009, p. 6) provided a more comprehensive definition of the concept that is twofold: exploratory search "can be used to describe an information-seeking problem context that is open-ended, persistent, and multi-faceted; and to describe information seeking processes that are opportunistic, iterative, and multi-tactical."The two aspects are not separable since the resolution of complex or vague information problems will definitely rely on non-linear search processes.

The Problem Context

Humans search because they realize the occurrence of information problems. In order to keep their lives and work running smoothly, they must deal with various tasks everyday, which provides the problem contexts for their search activities (Ingwersen and Järvelin 2005). Byström and Hansen (2005), Kim and Soergel (2005), and Li (2009), etc. have created different task classification frameworks. Tasks can be characterized based on many dimensions, but there are three essential and general ones, i.e. the specificity, volume, and timeliness of task goals (Marchionini 1995).

A highly specific task leads to the search of single facts, and users have the confidence to determine their validity. An unspecific task instead aims to engender interpretations or viewpoints, but users will be less certain about achieving their goals. Volume is reversely related to specificity. While a fact may be of low volume, containing merely a name, a number, or an image, interpretations or viewpoints usually need to be extracted from one or more documents. Timeliness refers to the expected time to acquire an answer. This can be as short as a moment or a few minutes, or as long as hours, days, or even months (Marchionini 1995).

In Marchionini (2006), exploratory tasks were distinguished from lookup tasks. The latter, also the basic kind of search tasks, involve discrete and well-structured information problems. That is, specific and finite search goals are immediately attainable. The former however become increasingly pervasive as both people's needs

and Web resources diversify. The seeking of information induced by ill-structured information problems is usually interwoven with learning or investigation. Searches that support learning or investigation aim to achieve the higher levels in Bloom's taxonomy of educational objectives.

The Search Process

A search process takes place within a particular problem context, and Wilson (1999) divided it into four stages: problem identification, problem definition, problem resolution, and solution statement. The transition from one stage to the next is always accompanied by the remarkable decrease of uncertainty. Uncertainty, a negative cognitive factor commonly seen in information seeking, will give rise to such affections as anxiety and lack of confidence (Kuhlthau 1999). In his communication theories, Shannon said that the more information people received, the lower their uncertainty. But in information science, it was thought that new information might sometimes result in the rebound of uncertainty especially during the earlier stages of the search process (Kalbach 2008).

The uncertainty aroused by exploratory problem context may fluctuate more evidently. Such fluctuation tends to ease as time progresses, with uncertainty decreasing meanwhile. But under some special circumstances, e.g. the search becoming more extensive and/or complex, it is possible that uncertainty will continue to fluctuate or even increase (White and Roth 2009). This can happen during any stage of the search process, and users will have to return to the previous stage so as to lessen the uncertainty again. As a result, an exploratory search process is made up of the four successive stages and the three feedback loops.

User behavior, unlike uncertainty, is the tangible and measurable variable in the search process. Wilson (1997), Choo et al. (2000), and Bates (2002) have investigated various information seeking modes. It is agreed that querying and browsing are the two basic active modes, i.e. users consciously investing time and energy to acquire information. While querying demands humans to recall from memory appropriate words to represent their information needs, browsing utilizes their perceptual abilities to recognize relevant information to their needs from the context (Marchionini 1995). The exploratory search process is characteristic of the alternation and iteration of the two modes (Marchionini 2006).

Theoretical Foundations of Exploratory Search Illustrated

Following the twofold definition of exploratory search, this study created two illustrations especially to ensure an easier and better understanding of the exploratory problem context and the exploratory search process. They will be further interpreted as follows.
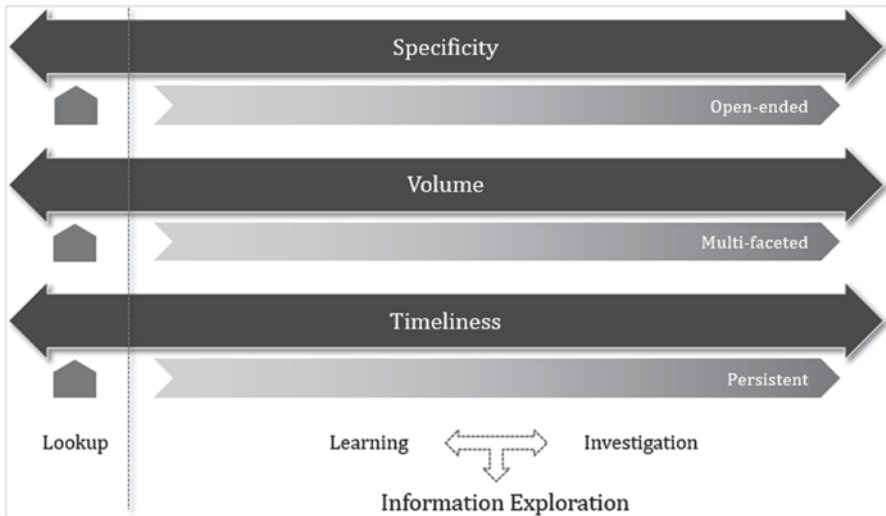
Fig. 1 The exploratory problem context model

Figure 1 represents each of Marchionini's (1995) dimensions of tasks, i.e. unspecificity, volume, and timeliness, with a continuum. With lookup problems being situated at the left ends on all three continua, exploratory problems occupy the remaining ranges. The less structured a problem is, the more cognitive resources users will have to invest, and the more closely the problem will approach the right ends on the continua where the characteristics of "open-ended", "multi-faceted", and "persistent" become the most significant.

Learning and investigation present two different levels of information exploration. Learning search is about accumulating existing knowledge on a certain topic or domain. What users anticipate are interpretive answers that help eliminate the unknown. The large volume of information objects they obtain can include texts, images, audios, and videos, etc. Some of these may verify or complement each other, but some may contradict or oppose each other. Users need to spend extra time viewing, comparing, and judging them therefore. Such internal cognitive processing activities will conduce towards a more solid human knowledge base. Investigative search, furthermore, is about creating new knowledge. Based on the analysis, synthesis, and assessment of the valuable contents extracted from information objects, users are capable of making intelligent decisions, planning, and predictions. This is a more advanced type of cognitive processing activity that usually lasts for a longer time and largely relies on users' current knowledge state to elicit evaluative answers embodying their own viewpoints.

In Fig. 2 a model of exploratory search process is presented. It adopts Wilson's (1999) four stages of the search process and integrates them with the behavioral characteristics of exploratory search. From the preliminary identification to sufficient definition of an information problem, users to a great extent rely on heuristic
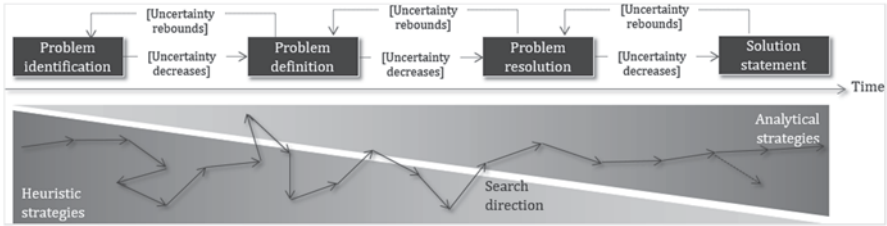
Fig. 2 The exploratory search process model

strategies in which browsing dominates. They navigate to potentially valuable collections of information and locate relevant concepts in the content via rapid scanning. It is important that they relate the concepts to one another to further clarify the core information need. When seeking answers to the problem, users instead adopt more frequently analytical strategies in which querying dominates. They decompose the information need into several parts that are more manageable and translate those parts into parallel or sequential queries. With the feedback from search systems users will gain a better understanding of the relevant concepts, which enables them to formulate more accurate queries and obtain more satisfying answers.

It should be noted that tentative querying before browsing is a component of heuristic strategies and targeted browsing after querying is a component of analytical strategies. Browsing is driven by external information, which gives users the opportunity to encounter new concepts of interest. The encountering will conduce to the generation of new needs and guide their searches to new directions. Internally driven querying seldom brings about an obvious change of the search direction. Nevertheless if a query returns no result, a new problem may be discovered. As a whole, users will proceed along an unpredictable non-linear path during the exploratory search process.

A Survey of Exploratory Search Systems

Thanks to the increasingly solid theoretical foundations, various exploratory search systems have been built to provide new technological capabilities and interface paradigms that facilitate the user-system interaction to improve the efficiency of querying and browsing (White et al. 2008). The support for query formulation and reformulation, in fact, is also very common in general search systems, such as query suggestion and expansion tools (Croft et al. 2010). However the support for search result browsing is exclusive in exploratory search systems. Existing systems have been trying to enhance users' abilities to understand and control massive result collections through information classification and visualization (White and Roth 2009).

Information Classification for Exploratory Search

As we know, mainstream Web search engines value precision, especially the high relevance of the results on the first search result page. Differently, exploratory search systems pay more attention to recall because the lower-ranking pages may also contain useful information (Marchionini 2006). It is thus necessary to relieve users' browsing burden when they navigate through each page. Exploratory search systems have introduced a variety of methods to classify search results for this purpose. With many results divided into a few groups, users are more able to identify the key information (Jiang and Koshman 2008).

Classification is a process that involves "systematic arrangement of entities based on analysis of the set of individually necessary and jointly sufficient characteristics that defines each class" (Jacob 2004). This study conducted a comprehensive survey on the result classification methods employed by exploratory search systems, including fully functional systems that are/were available to ordinary users as well as prototype systems mentioned in the literature. As indicated by the survey, there are four major ways to decrease the density of the result space, i.e. hierarchical classification, faceted classification, dynamic clustering, and social classification.

Hierarchical Classification

Hierarchical classification refers to a system of fixed non-overlapping classes within a hierarchical enumerative structure to exactly reflect a pre-determined ordering of reality. It results from the top-down division of the information space according to some "logic" (Taylor and Wynar 2004). The parent-child relationships between superordinate and subordinate classes are usually presented in trees. The use of general hierarchical classification systems, e.g. Dewey Decimal Classification and Library of Congress Classification, to arrange library resources can be traced back to the 19th century. Nowadays, Yahoo! Directory and Open Directory Project are the two most widely known hierarchical classification systems for Web resources. They are compiled and maintained by experts and users respectively.

Hierarchical classification has been used to organize search results in several studies. Chen and Dumais (2000) developed an interface where webpages returned by the search engine were assigned into the classes of LookSmart, a Web directory, on the fly with text classification algorithms. They found that users were 50 % more efficient at finding information on this category interface than on the list interface. CitiViz was a visual search interface that displayed an overview of the document sets in a digital library based on the ACMComputing Classification System. Its effectiveness exceeded the traditional list in various exploratory tasks (Kampanya et al. 2004). Besides, hierarchical classification can help improve the internal search of websites. For instance, the website of UC Berkeley once introduced the Cha-Cha system that showed within-site search results in its own hierarchical sitemap (Chen

et al. 1999). Another similar example is the WebTOC system by the HCI Lab at the University of Maryland (Nation 1998).

These are early attempts to create exploratory search systems and they have a common preference for hierarchical classification to enhance search result organization. Provided with a familiar and stable hierarchical classification, users are able to establish their mental models about the whole result space rapidly and to see their positions in the space. On the one hand, their familiarity with the classification system can reduce the difficulties in grasping the system. On the other hand, the stableness of the classification system can lessen their anxiety in the search process. Nevertheless it is not easy to make efficient use of hierarchical classification in result organization. One thing to consider is how to balance the breadth and depth of the hierarchy. Also the problem of polyhierarchy (i.e. an item falling into two different categories at the same time) needs to be addressed (Morville and Rosenfeld 2006).

Faceted Classification

Faceted classification, simply speaking, is composed of several facets and a number of categories under each facet (Tunkelang 2009).The facet corresponds to an attribute of the information collection and the categories contained represent various values of that attribute (Hearst 2006). As early as the 1960s, Ranganathan (1960) introduced the notion "facet" to library and information science. In his Colon Classification Scheme, the five fundamental facets are personality, matter, energy, space, and time. But in most cases facets are created for particular domains, such as the author, language, and year of a book, or the price, brand, and size of a laptop.

Flamenco (Hearst 2006), mSpace (Schraefel et al. 2005), and Relation Browser (Capra and Marchionini 2008) are pioneer studies which applied faceted classification in search. These prototype systems, though different in terms of information type and interface design, all provide a set of small categorical hierarchies instead of one large cover-all topical hierarchy. Users are allowed to browse the hierarchies one by one and select the most appropriate category in each, which enables them to narrow down the search scope gradually. Related user studies showed that faceted classification was easy to understand, and many searchers preferred this approach for it avoided empty results and supported exploration and discovery (Yee et al. 2003).

We can find faceted classification in a wide variety of search environments. On E-commerce platforms, both C2 C (e.g. eBay and Taobao) and B2 C (e.g. Overstock and Bestbuy), faceted search are making full use of products' structured metadata to improve their find ability, producing great business value (Dash et al. 2008). In addition, next-generation library catalogs are now featuring faceted search. Many university libraries, such as those of Duke University, Harvard University, and the University of Pittsburgh, depend on discovery service providers (e.g. Endeca, AquaBrowser, and Summon) to offer faceted browsing experience to their patrons (Yang and Wagner 2010).

By taking multiple conceptual dimensions into consideration, faceted classification better satisfies different users who view the world differently. It is an effective way to cope with the challenges in information organization brought about by compound concepts. And faceted search is in essence a form of exploratory search. After the search results are mapped onto a faceted classification system, users can look into them in a more flexible manner, i.e. examining any number of facets in any order. If combining the labels of all the categories ever selected, one can see a complex Boolean query. This approach favors recognition over recall to alleviate human mental work. Thanks to the logical and predictable structure of faceted classification, faceted search systems will become the prevailing search tools in electronic environments.

Dynamic Clustering

The basic idea of clustering is grouping information items by algorithms so that the items within one group are similar or relevant and different groups are obviously distinct (Manning et al. 2008). Since van Rijsbergen's (1979) Cluster Hypothesis – "closely associated documents tend to be relevant to the same request", more and more researchers in the area of information retrieval deemed clustering the retrieved documents into groups with common subjects a natural alternative to ranking them in a linear list (Croft and Leouski 1996).

Vivisimo Enterprise Search was among the first clustering search systems in practice. It was characteristic of post-retrieval clustering, a three-step process: (1) generating the clustering structure based on the content of the search results; (2) inserting the result items into appropriate categories in the structure; and (3) selecting and preparing the categories to be presented to users (Koshman et al. 2006). Clusty (now Yippy), one of the most influential clustering search engines on the Web, was built upon the technical support of Vivisimo. Other leading systems include iBoogie, PolyMeta, and Carrot2, etc., but some early systems, such as Grokker, KartOO, WebClust, and Mooter, have been shut down for various reasons. These systems mostly perform clustering on the top search results and their clustering structures can be single-level or multi-level (Jiang and Koshman 2008).

The usability of a clustering structure is largely determined by the quality of category label. Carpineto et al. (2009) divided clustering algorithms according to their category description methods into three types, i.e. data-centric, description-aware, and description-centric. Clustering search engines often adopt the description-centric algorithms. They emphasize that the descriptions of category labels should be simple and clear and that undescribable categories should be removed for being of little value to users. In general, clustering search engines will also support metasearch. More specifically, they obtain and aggregate search results from Google, Bing, and other Web search engines via API and instead focus on the clustering work. Metasaearch compensates for the limited scope of a single search engine index, which helps users achieve the comprehensive examination of search results on a uniform interface (Morville and Callender 2010).

Clustering technologies are of great significance to exploratory search. The best of clustering is that the classification structure is automatically generated for the current situation. Dynamic classification gets rid of the complexity and cost of building and maintaining a fixed scheme. In addition to providing users a convenient way to view the results under specific topics, clustering solves the problem of polysemy. The results are differentiated according to their meanings, facilitating users to make selective browsing. Furthermore, clustering gathers the related results that originally scatter on different search result pages. With all the important topics surfacing at once, users can review the whole result space in a more systematic manner.

Social Classification

Social classification, also known as folksonomy, is made up of people-contributed free tags and takes the form of a flat and loose namespace (Kroski 2005). This type of classification is firstly seen in social tagging systems where users assign tags to resources for the purpose of self-organization (Smith 2007). Depending on the tagging privilege, it can be narrow or broad (Golder and Huberman 2006). Flickr, Vimeo, Reddit, and LiveJournal etc. are representative social tagging systems featuring narrow folksonomies, while BibSonomy, Folkd, LibraryThing, and Douban etc. broad folksonomies. In these systems, users tend to explore the resources that have already been tagged by others (Millen and Feinberg 2006). Usually, users who are accustomed to discovering resources by tag are active tag contributors. Since tags express explicit topics, they can increase the directedness of the browsing process as intermediaries (Jiang 2013).

Amazon, a diversified E-commerce platform, has introduced product tagging. When looking for products, customers may conduct tag search, i.e. the query being recognized as a tag. All the products to which the tag has been assigned will be returned, and the suggestions of relevant tags allow users to refine the results further. In Amazon, social classification is independent of the existing hierarchical departments of products. Similarly, the libraries of the University of Pennsylvania and the University of Michigan also have complemented their traditional hierarchical classification of book resources with social classification, engendering the PennTags and Mtagger systems respectively (Pirmann 2012).

As a basic classification method on the Web 2.0 and a supplemental method on the Web 1.0, social classification shows potential in exploratory search for being inexpensive to create and responsive to changes. Tagging is essentially an individual activity because people tag according to their personal understanding and in a distributed manner. However the social aspect of tagging consists in the fact that tags are aggregated by the system. At the micro level, the bibliographic record of each resource is composed of the tags ever attached to it; and at the macro level, all the tags from all the users constitute a classification system. When users tag a resource, they not only facilitate their own future retrieval of the resource, but also create a path for others to find it.

Information Visualizationfor Exploratory Search

Many exploratory search systems provide visualization tools to aid the presentation of search results after they are classified or grouped. Simply speaking, visualization is showing abstract information with intuitive graphs. There are three elements in Spence's (2007) visualization process model: representation, presentation, and interaction. Visualizations represent data values and relations in various forms, present them in constrained spaces, and allow users to select the required view via interaction. Since interaction is in the control of users, their perception and cognition have a strong impact on the effectiveness of visualizations (Tory and Moller 2004). Human's perceptual system is responsible for importing the representations, and cognitive system adding meaning to them and storing the consequent understanding in memory (Spence 2001).

As Koshman (2006, p. 20) pointed out, "the notion of visualization supporting exploratory search can be an extremely powerful model that applies the high bandwidth of human perceptual processing to reduce or mediate uncertainty surrounding initial queries and to see new relationships among the retrieved data set that would not be present in a traditional linear search result listing." It was noticed in the survey of exploratory search systems that each of the above ways of search result classification had aroused some interest in the design and development of corresponding visualizations.

Visualizations for Hierarchical Classification

Given its inherent structural traits, hierarchical classification is often associated with the tree visualization. A representative example is the CitiViz search interface already mentioned (Fox et al. 2006). In addition to an expandable tree list (Fig. 3 left), it introduced a hyperbolic tree (Fig. 3 upper right) and a 2D scatter plot (Fig. 3 middle right). Hyperbolic trees are generated by misshaping the original tree structure. The distortion will enlarge the branches of interest with more details and meanwhile shrink the adjacent branches to occupy less space, supporting the "focus+context" display (Lamping et al. 1995).This hyperbolic tree consists of rectangle nodes and bubbles attached to them. They respectively represent subject categories and the result document sets falling into the categories. A single click on a node will bring it from context to focus smoothly. The size of each bubble is proportional to the quantity of documents in it. When a bubble is selected, the documents contained will map onto the scatter plot where the x-axis is rank and the y-axis date. The towers on the scatterplot stand for individual documents with the layer colors indicating the subject categories which the documents belong to. CitiViz color-coded the topical categories and used the coding system to connect three different visualization views. It not only catered to different users' perceptual habits but also reinforced their understanding with multiple levels of details.
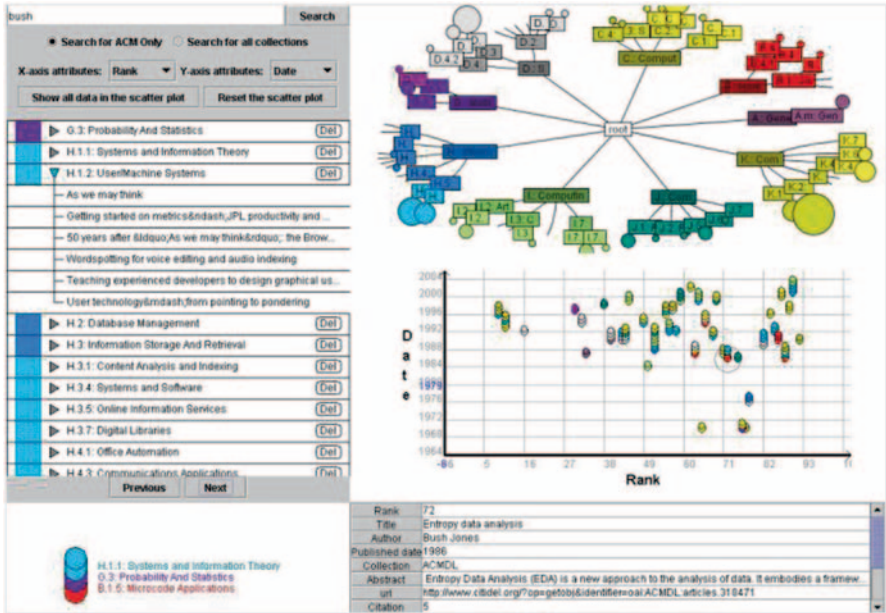
Fig. 3 The CitiViz search interface

Also worth mentioning is ResultMap designed by Clarkson et al. (2009), a search tool based on the treemap visualization. Treemaps transform tree structures into recursively nested rectangle zones, making good use of space. Each rectangle is filled with smaller rectangles, indicating the parent-child relationships. The area of a rectangle is often in proportion to the value of a particular attribute describing the dataset (Shneiderman and Wattenberg 2001). As shown in Fig. 4, Result Map demonstrated all the documents in a knowledge repository on a treemap according to their hierarchical relationships and ensured a stable expression of the entire information space. The result documents returned by each query will be highlighted on the treemap and the colors suggested their types so that users can access the details of the documents. The treemap appears on every search result page, right beside the result list. In particular, mouse hover on a certain rectangle will change the display of related results in the list and vice versa. The interaction between the visual and textual presentations is therefore made possible.

Visualizations for Faceted Classification

Most faceted search systems, strictly speaking, are actually text-based. For example, Flamenco just distinguished the facets with colors. It is perhaps because the textual interfaces are already easy to understand and use, not much energy
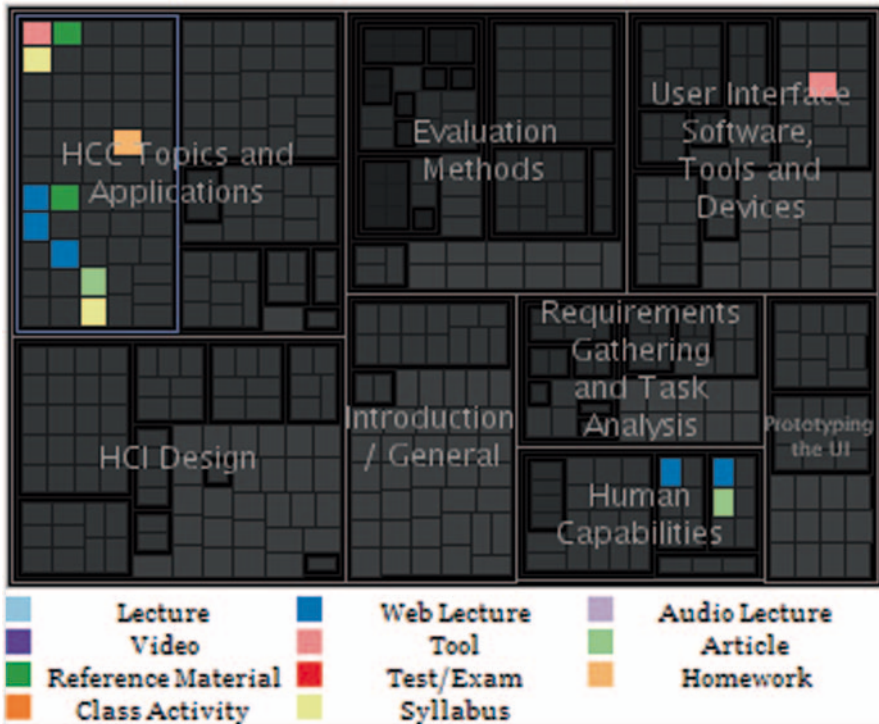
Fig. 4 The result maptreemap visualization

has been devoted to developing visualizations for faceted classification. The most remarkable attempt so far should be FacetMap by Smith et al. (2006).This purely graphic system employed round-cornered rectangles and ovals to represent facets and their categories respectively, as seen in Fig. 5a. More frequently used facets will appear larger on the screen with more categories exposed, but all the ovals are of the same size with the exact numbers of items contained provided under the category labels. Users can easily drill down to the information items at the lowest level by selecting relevant facets, categories, and sub-categories along the way (Fig. 5b). In fact FacetMap realized the "overview+detail" display that was different from distortion. When a facet is enlarged to show more details through semantic zooming, other facets are excluded from the limited screen. Users may lose the control of interaction and even feel disoriented (Heo and Hirtle 2001).

Visualizations for Dynamic Clustering

Unlikely, visualizations are a common component of clustering search systems. Although text-based tree lists are widely used, visualizations are able to reveal
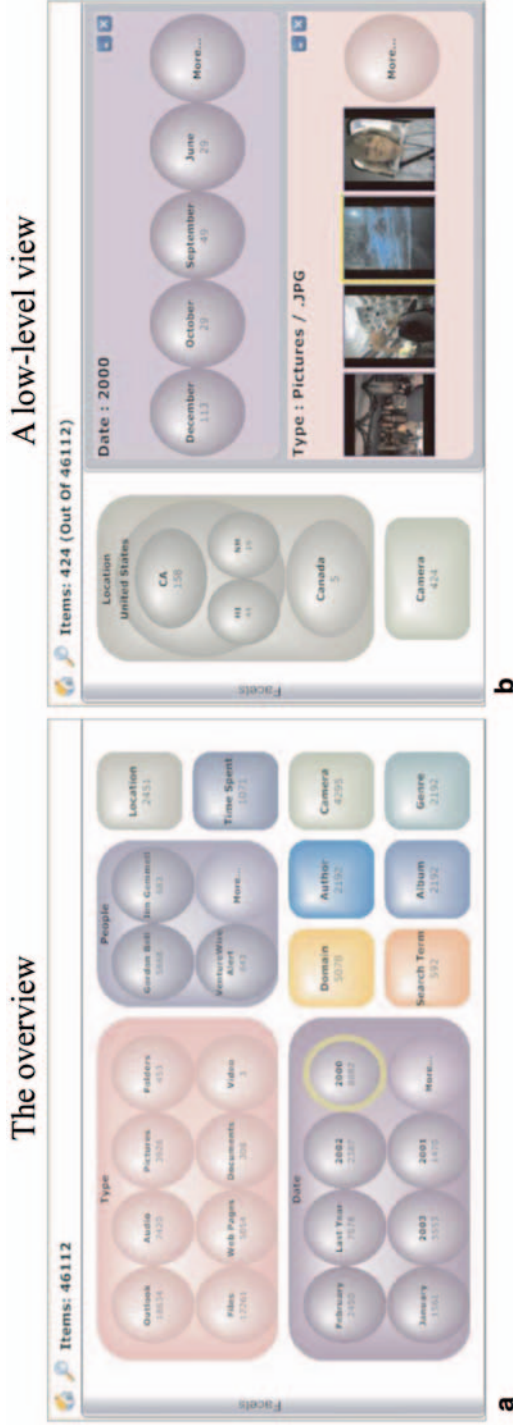
Fig. 5 The FacetMap visualization for faceted search. (**a**) The overview. (**b**) A low-level view
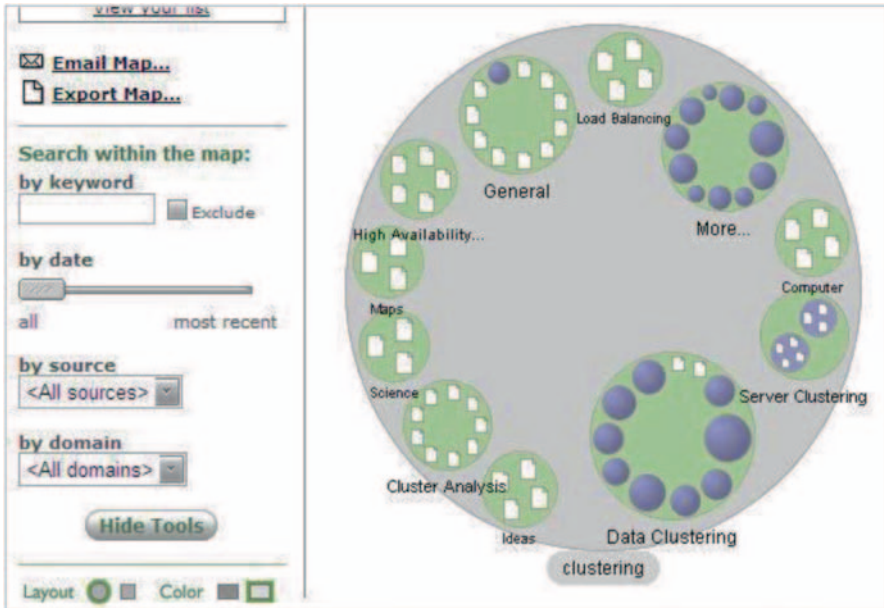
Fig. 6 Grokker's map view

the relationships between clusters and items more efficiently for possessing richer spatial attributes. The abovementioned Grokker, KartOO, and Carrot2 have developed interactive 2D visualizations that facilitated the examination of search results (Koshman 2006; Kothari 2010). Grokker's map view (Fig. 6) followed the "overview+detail" display to show the nesting of categories (green circles), subcategories (blue circles), and result items (white page icons), and users were supported to move forward or trace back level by level. KartOO positioned result items (yellow document icons) within the same cluster on a cartographic map (Fig. 7). One can see the connections between adjacent items, and the labels in between indicate the subject they share. Carrot2 offers two visualization views, i.e. Circles (Fig. 8a) and Foam Tree (Fig. 8b), which differ in shape. The colored zones representing the clusters are arranged by cluster size.

3D approaches involving real-world metaphors have been proposed to visualize clustered results. Figure 9 shows a prototype visualization module that presents the search results from Carrot2 in a new way (Akhavi et al. 2007). The algorithm traverses the original clustering hierarchy and transforms the clusters into tree branches and result items fruits in a 3D space. Bonnel et al. (2006), innovatively, used the metaphor of cities. Result items are visualized as buildings, with the neighboring districts standing for related topics (Fig. 10). Building height suggests result relevance and building surface is filled with the page snapshot of the result. 3D visualizations, however, were thought to be ineffective because the third dimension could inhibit users and make the interface more confusing (Risden et al. 2000).

Fig. 7 KartOO' cartographic map

What's more, displaying 3D visualizations on 2D devices is in itself problematic (Modjeska 2000).

Visualizations for Social Classification

The tag cloud visualization came into being to address the structural looseness of social classification. It is a text-based visualization method that displays the tags in alphabetical order and indicates their frequencies with font size. Most tag clouds only include the most active tags for they reflect the popular topics people are concerned with recently (Sinclair and Cardew-Hall 2008). One will be redirected to all the resources associated with a specific tag by a simple click on that tag; and sometimes, the click may also lead to the users who have added the tag and/or other co-assigned tags. The insufficiencies of the tag cloud are also obvious, and a major one is that semantically related tags may scatter in the cloud because they are not alphabetically close. The efficiency of a cloud will be greatly influenced when it reaches a certain scale. It is difficult for users to quickly identify the most useful ones from tens of thousands of tags (Hearst and Rosner 2008).

Researchers have been improving tag clouds. In Hassan-Montero and Herrero-Solana (2006), insignificant tags (e.g. "toread" and "diy") were removed from the cloud and synonymies were merged to make space for more substantial tags (e.g. "philosophy" and "religion"). After lowering the semantic density, the researchers
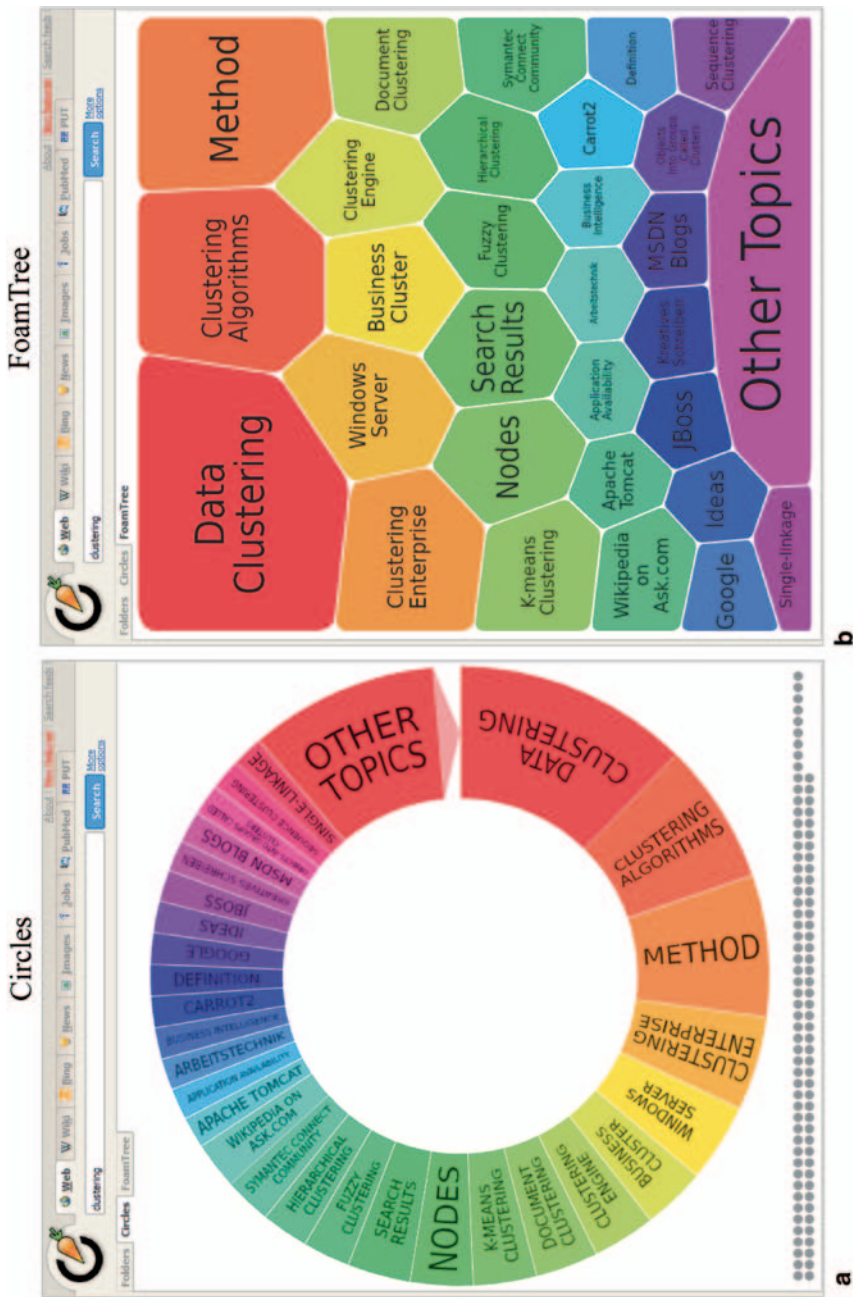
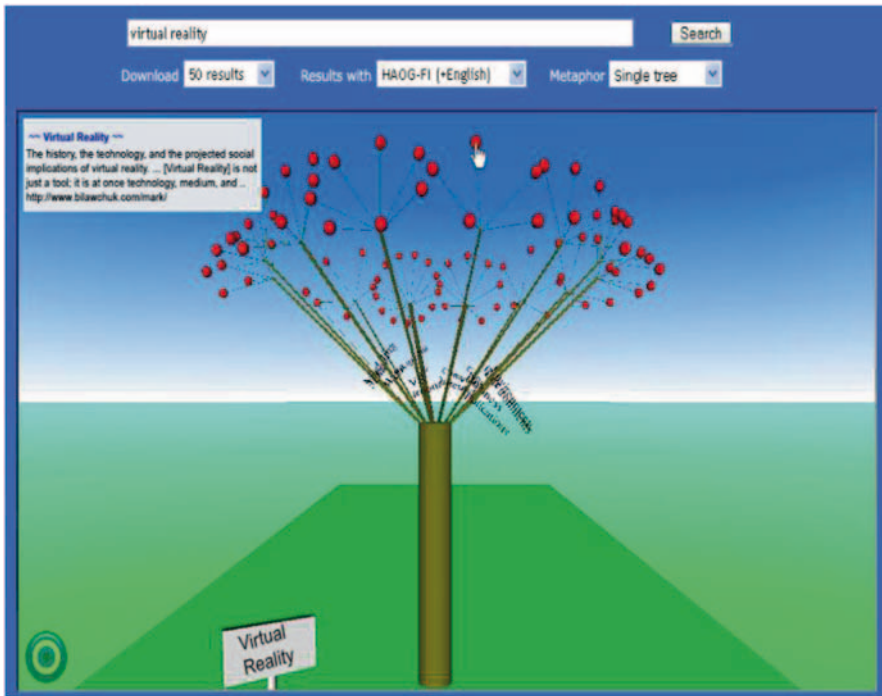**Fig. 8** Carrot2's visualization views. (**a**) Circles. (**b**) FoamTree

Fig. 9 3D tree visualization for clustering search

changed the layout of the tag cloud with clustering algorithms: frequently co-occurrent tags appear on the same row (Fig. 11). This is conducive to topic differentiation and knowledge discovery. Chen et al. (2010) created the TagClusters visualization, a tag cloud variation based on tag clustering. In this brand new view (Fig. 12), tags are no longer displayed in rows; instead, their relative positions are determined by co-occurrence. Semantically related tags determined by text analysis will form a tag group as represented with the translucent pink zone. The name of a group, i.e. the purple uppercase label, is in proportion to the total frequencies of all the tags in that group. A tag group may further contain sub-groups, and different sub-groups can overlap. This view facilitates users to understand the affiliations and associations between tags.

The Future of Exploratory Search

There is no denying that the technological development in information classification and visualization is an impetus to exploratory search systems. A handful of researchers however have recognized that the future of exploratory search lied in
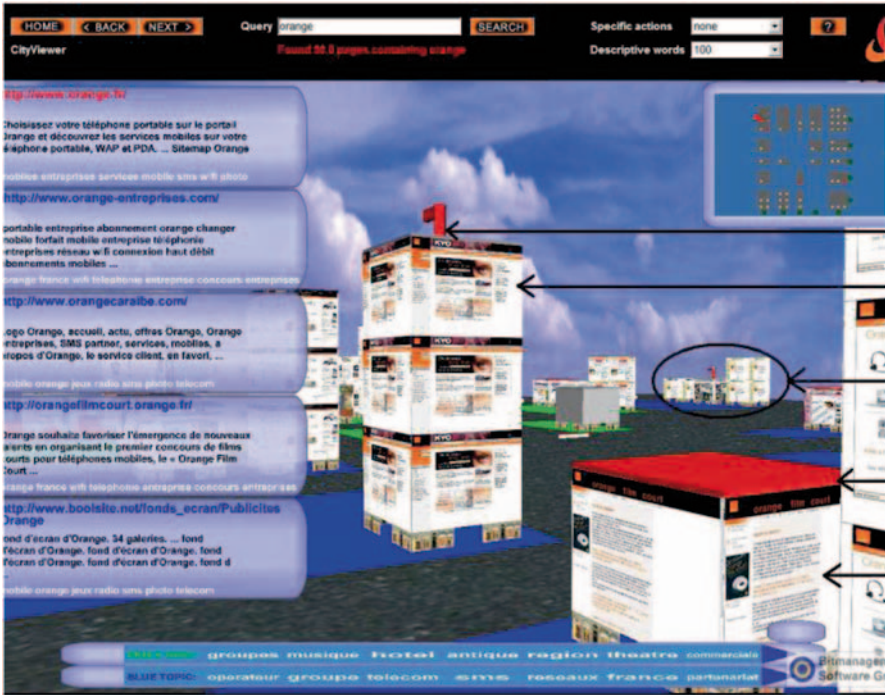
Fig. 10 3D city visualization for clustering search

Fig. 11 A clustering-based tag cloud

Fig. 12 TagClusters

the vast social space. Evans and Chi (2008) found based on a survey of 150 participants that interpersonal communication played an indispensable role throughout the entire search process, including the pre-search problem statement, information collecting and selecting, and post-search result sharing. In Kammerer et al. (2009), tag data from a social bookmarking site was added to search results and user feedback was used to further improve the relevance of result listings. The experiment suggested that exploration of new knowledge in ill-structured domains could be effectively supported in this way.

Social interaction, both explicit and implicit, will become a core component of exploratory search in the near future. People are not separated from one another during information seeking. They may acquire information from others out of various reasons, and such tendency can be very strong (Chi 2009). Morville and Rosenfeld (2006) also deemed seeking help from others an information seeking mode as important as querying and browsing. In existing exploratory search systems, nevertheless, users are still independent searchers in the traditional sense even though their exploration activities have become more effective with system-offered informational clues.

In the Web 2.0 era, the growth of social software has brought about wider and more frequent communication and sharing of information. People's everyday information seeking is inevitably mixed with their social interaction, which will create new possibilities for exploratory search systems. One the one hand, human-to-human conversations are beneficial to lowering vocabulary barriers. Querying in more natural ways will reduce users' cognitive loads. On the other hand, the "collective intelligence" of many individuals can produce social clues. In other words, new comers may follow the trails of actions left by previous users to identify appropriate

browsing paths already taken by the majority. Svensson (1998) distinguished these two types of social interaction as direct and indirect social navigation.

Navigation is searching without a clear goal, and social navigation is navigation guided by human beings (Svensson 2002). Direct social navigation means that navigators seek personalized advice from others through two-way communication. In this way they may not only find the answers to such basic questions as "where am I", but also stand a chance of clarifying their goals and choosing a correct path towards the destination. Indirect social navigation, in contrast, features one-way communication in which advice givers provide guidance to navigators unintentionally. This takes the form of "cumulative information", a dynamic concept. People entering and occupying the information space break its original design and influence its growth, just like that the regularly walked track in the forest becomes a road (Svensson 1998).

In the early days social navigation support systems were mostly history-enriched environments on the basis of indirect social navigation. The rise of social software since 2005 provides a promising setting of research. Millen and Feinberg (2006) found in a study on the social bookmarking service dogear that viewing others' bookmark collections and clicking on tags to view the associated bookmarks were the commonest forms of social navigation. Vosinakis and Papadakis (2011) integrated spatial, semantic, and social navigation in the 3D environments of virtual worlds. The prototype framework they proposed included thematic discussions, user trails and tags, semantic filters, linked data and other features. Shami (2011) designed a social file sharing system, Cattail. It supported social navigation through a recent events stream and downloading history sharing. System evaluation results implied that Cattail could help users discover more relevant people and content.

In summary, the existing research on exploratory search has been focusing on individual users' search activities, ignoring the significance of social support to information exploration. There is a natural trend that social navigation research merges into this area. We may gain a great deal of enlightenment from the findings on both direct and indirect social navigation. The boom of social software, at the same time, increases the feasibility of realizing social interaction in exploratory search. Others' advice or activities usually have a strong impact on people's informational decisions. The interest in social interaction will diversify future research on exploratory search.

References

Akhavi MS, Rahmati M, Amini NN (August 2007). 3d visualization of hierarchical clustered web search results. In Computer Graphics, Imaging and Visualisation, 2007. CGIV'07. IEEE. pp 441–446

Bates MJ (1989) The design of browsing and berrypicking techniques for the online search interface. Online Inf Rev 13(5):407–424

Bates MJ (2002) Toward an integrated model of information seeking and searching. New Rev Inf Behav Res 3:1–15

Belkin NJ (1996) Intelligent information retrieval: whose intelligence. ISI 96:25–31

Bonnel N, Lemaire V, Alexandre CH, Morin A (2006, February) Effective organization and visualization of web search results. In IASTED International Conference on Internet and Multimedia Systems and Applications (EuroIMSA'06). pp 209–216

Byström K, Hansen P (2005) Conceptual framework for tasks in information studies. J Am Soc Inf Sci Technol 56(10):1050–1061

Capra RG, Marchionini G (2008, June) The relation browser tool for faceted exploratory search. In Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries. ACM. pp 420–420

Carpineto C, Osiński S, Romano G, Weiss D (2009) A survey of web clustering engines. ACM Comput Surv (CSUR) 41(3):17

Chen H, Dumais S (2000, April) Bringing order to the web: automatically categorizing search results. In Proceedings of the SIGCHI conference on Human factors in computing systems. ACM. pp 145–152

Chen M, Hearst M, Hong J, Lin J (1999) Cha-Cha: a system for organizing intranet search results. In Proceedings of the 2nd USENIX Symposium on Internet Technologies and Systems pp 1–14

Chen YX, Santamaría R, Butz A, Therón R (2010) TagClusters: Enhancing semantic understanding of collaborative tags. Int J Creat Interfaces Comput Graph (IJCICG) 1(2):15–28

Chi EH (2009) Information seeking can be social. Computer 42(3):42–46

Choo CW, Detlor B, Turnbull D (2000) Information seeking on the web: an integrated model of browsing and searching. First Monday 5(2)

Clarkson E, Desai K, Foley J (2009) Resultmaps: visualization for search interfaces. Vis Compu Graph IEEE Trans 15(6):1057–1064

Croft WB, Leouski AV (1996) An evaluation of techniques for clustering search results. Computer Science Department Faculty Publication Series, 36

Croft WB, Metzler D, Strohman T (2010) Search engines: information retrieval in practice. Reading, Addison-Wesley, p 283

Dash D, Rao J, Megiddo N, Ailamaki A, Lohman G (2008, October) Dynamic faceted search for discovery-driven analysis. In Proceedings of the 17th ACM conference on Information and knowledge management. ACM. pp 3–12

Evans BM, Chi EH (2008, November) Towards a model of understanding social search. In Proceedings of the 2008 ACM conference on Computer supported cooperative work. ACM. pp 485–494

Fox EA, Neves FD, Yu X, Shen R, Kim S, Fan W (2006) Exploring the computing literature with visualization and stepping stones pathways. Commun ACM 49(4):52–58

Golder SA, Huberman BA (2006) Usage patterns of collaborative tagging systems. J Inf Sci 32(2):198–208

Hassan-Montero Y, Herrero-Solana V (2006, October) Improving tag-clouds as visual information retrieval interfaces. In International Conference on Multidisciplinary Information Sciences and Technologies, pp 25–28

Hearst MA (2006) Clustering versus faceted categories for information exploration. Commun ACM 49(4):59–61

Hearst MA, Rosner D (2008, January) Tag clouds: data analysis tool or social signaller? In Hawaii International Conference on System Sciences, Proceedings of the 41st Annual. IEEE. pp 160–160

Heo M, Hirtle SC (2001) An empirical comparison of visualization tools to assist information retrieval on the Web. J Am Soc Inf Sci Technol 52(8):666–675

Ingwersen P (1996) Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. J Doc 52(1):3–50

Ingwersen P, Järvelin K (2005) The turn: integration of information seeking and retrieval in context, vol 18. Springer

Jacob EK (2004) Classification and categorization: a difference that makes a difference. Libr Trends 52(3):515–540

Jiang T (2013) An exploratory study on social library system users' information seeking modes. J Doc 69(1):6–26

Jiang T, Koshman S (2008) Exploratory search in different information architectures. Bull Am Soc Inf Sci Technol 34(6):11–13

Kalbach J (2008) On uncertainty in information architecture. J Inf Archit 1(1):48–56

Kammerer Y, Nairn R, Pirolli P, Chi EH (2009, April) Signpost from the masses: learning effects in an exploratory social tag search browser. In Proceedings of the SIGCHI conference on human factors in computing systems. ACM. pp 625–634

Kampanya N, Shen R, Kim S, North C, Fox EA (2004) CitiViz: a visual user interface to the CITIDEL system. In research and advanced technology for digital libraries. Springer, Berlin, pp 122–133

Kim S, Soergel D (2005) Selecting and measuring task characteristics as independent variables. Proc Am Soc Inf Sci Technol 42(1)

Koshman S (2006) Exploratory search visualization: identifying factors affecting evaluation. EESS 2006, 20

Koshman S (2006) Visualization-based information retrieval on the web. Libr Inf Sci Res 28(2):192–207

Koshman S, Spink A, Jansen BJ (2006) Web searching on the vivisimo search engine. J Am Soc Inf Sci Technol 57(14):1875–1887

Kothari SS (2010) Evaluating the efficacy of clustered visualization in exploratory search tasks. Purdue University, West Lafayette, Indiana

Kroski E (2005) The hive mind: Folksonomies and user-based tagging. Library 2:91–103

Kuhlthau CC (1999) The role of experience in the information search process of an early career information worker: perceptions of uncertainty, complexity construction, and sources. J Am Soc Inf Sci 50(5):399–412

Lamping J, Rao R, Pirolli P (1995, May) A focus+ context technique based on hyperbolic geometry for visualizing large hierarchies. In Proceedings of the SIGCHI conference on Human factors in computing systems. ACM Press/Addison-Wesley Publishing Co., pp 401–408

Li Y (2009) Exploring the relationships between work task and search task in information search. J Am Soc Inf Sci Technol 60(2):275–291

Manning CD, Raghavan P, Schütze H (2008) Introduction to information retrieval, vol 1. Cambridge University Press, Cambridge

Marchionini G (1995) Information seeking in electronic environments. Cambridge University Press, Cambridge

Marchionini G (2006) Exploratory search: from finding to understanding. Commun ACM 49(4):41–46

Millen DR, Feinberg J (2006, June) Using social tagging to improve social navigation. In Workshop on the Social Navigation and Community based Adaptation Technologies

Modjeska DK (2000) Hierarchical data visualization in desktop virtual reality (Doctoral dissertation, University of Toronto)

Morville P, Callender J (2010) Search patterns. O'Reilly

Morville P, Rosenfeld L (2006) Information architecture for the World Wide Web: designing large-scale web sites. O'Reilly Media

Nation DA (1998, April) WebTOC: a tool to visualize and quantify Web sites using a hierarchical table of contents browser. In CHI 98 Cconference Summary on Human Factors in Computing Systems. ACM. pp 185–186

Nolan M (2008) Exploring exploratory search. Bull Am Soc Inf Sci Technol 34(4):38–41

Pirmann C (2012) Tags in the catalogue: insights from a usability study of librarything for libraries. Libr Trends 61(1):234–247

Pirolli P, Card S (1999) Information foraging. Psychol Rev 106(4):643–675

Ranganathan SR (1960) Colon classification: basic classification, 6th edn. Asia Publishing House, New York

Risden K, Czerwinski MP, Munzner T, Cook DB (2000) Initial examination of ease of use for 2D and 3D information visualization of Web content. Int J Hum-Comput Stud 53(5):695–714

Saracevic T (1996) Modeling interaction in information retrieval (IR): a review and proposal. In Proceedings of the ASIS annual meeting, vol 33. pp 3–9

Saracevic T (1997, January) The stratified model of information retrieval interaction: extension and applications. In Proceedings of the ASIS annual meeting, vol 34. pp 313–327

Schraefel MC, Alex D, Smith E, Russel A, Owens A, Harris C, Wilson M (2005) The mSpace classical music explorer: improving access to classical music for real people. In MusicNetwork Open Workshop, Integration of Music in Multimedia Applications

Shami NS, Muller M, Millen D (2011) Browse and discover: social file sharing in the enterprise. In Proceeding of the ACM 2011 Conference on Computer Supported Cooperative Work, pp 295–304

Shneiderman B, Wattenberg M (2001, October) Ordered treemap layouts. In Proceedings of the IEEE Symposium on Information Visualization 2001, vol 73078

Sinclair J, Cardew-Hall M (2008) The folksonomy tag cloud: when is it useful? J Inf Sci 34(1):15–29

Smith G (2007) Tagging: people-powered metadata for the social web, safari. New Riders

Smith G, Czerwinski M, Meyers BR, Robertson G, Tan DS (2006) FacetMap: a scalable search and browse visualization. Vis Comput Graph IEEE Trans 12(5):797–804

Spence R (2001) Information visualization, 1st edn. Addison-Wesley

Spence R (2007) Information visualization: design for interaction, 2nd edn. Prentice Hall

Spink A (1997) Study of interactive feedback during mediated information retrieval. J Am Soc Inf Sci 48(5):382–394

Svensson M (1998) Social navigation, in Dahlback N. Exploring navigation: towards a framework for design and evaluation of navigationin electronic spaces. Swedish Institute of Computer Science, pp 73–88

Svensson M (2002) Defining, designing and evaluating social navigation (Doctoral dissertation, Stockholm University)

Taylor AG, Wynar BS (2004) Wynar's introduction to cataloging and classification. Libraries Unlimited Inc.

Tory M, Moller T (2004) Human factors in visualization research. IEEE Trans Vis Comput Graph 10(1):72–84

Tunkelang D (2009) Faceted search. Synth Lect Inf Concepts Retr Serv 1(1):1–80

van Rijsbergen CJ (1979) Information retrieval. Butterworths, London

Vosinakis S, Papadakis I (2011) Virtual worlds as information spaces: supporting semantic and social navigation in a shared 3D environment. In Proceedings of 3rd International Conference on Games and Virtual Worlds for Serious Applicants, pp. 220–227

White RW, Kules B, Bederson B (2005, December) Exploratory search interfaces: categorization, clustering and beyond: report on the XSI 2005 workshop at the Human-Computer Interaction Laboratory, University of Maryland. In ACM SIGIR Forum, vol 39, No. 2. ACM. pp 52–56

White RW, Marchionini G, Muresan G (2008) Evaluating exploratory search systems: introduction to special topic issue of information processing and management. Inf Proc Manag 44(2):433–436

White RW, Roth RA (2009) Exploratory search: beyond the query-response paradigm. Synth Lect Inf Concept Retr Serv 1(1):1–98

Wilson TD (1997) Information behaviour: an interdisciplinary perspective. Inf Process Manag 33(4):551–572

Wilson TD (1999) Models in information behaviour research. J Doc 55(3):249–270

Yang SQ, Wagner K (2010) Evaluating and comparing discovery tools: how close are we towards next generation catalog? Libr Hi Tech 28(4):690–709

Yee KP, Swearingen K, Li K, Hearst M (2003, April) Faceted metadata for image search and browsing. In Proceedings of the SIGCHI conference on Human factors in computing systems. ACM. pp 401–408