Introduction to Special Issue of IP&M on Human-Computer Information Retrieval

Ryen W. White
Microsoft Research
Redmond, WA 98052 USA
ryenw@microsoft.com

Bill Kules
The Catholic University of America
Washington, DC 20064 USA
kules@cua.edu

Rob Capra
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599 USA
rcapra@unc.edu

Catherine Smith
Kent State University
Kent, OH 44242 USA
csmit141@kent.edu

Gene Golovchinsky
FX Palo Alto Laboratory
Palo Alto, CA 94304 USA
gene@fxpal.com

Daniel Tunkelang
LinkedIn
Mountain View, CA 94043 USA
dtunkelang@linkedin.com

Contemporary search engines are optimized for look-up scenarios where the information target is well-defined and human-machine interaction is limited to queries and search-result selections. In this role the search system serves as a cognitive prosthetic, temporarily enhancing people's mental capabilities to provide access to additional information not known to the searcher or not readily accessible to them. However, this type support is insufficient for tasks requiring more involved information interaction (e.g., where cross-query learning may be important) and situations where information behavior encompasses more than just information *seeking*. People may often possess more aspirational goals such as augmenting their intellect (Englebart, 1962) and applying their knowledge to explore, learn, and make sense of the information they encounter (Marchionini, 2006a). To meet these demands search systems need to form a symbiotic relationship with their users, providing support for fluid and meaningful interaction between human and machine, and promoting critical thinking and increased interaction with information to facilitate cognitive development.

Human-computer information retrieval (HCIR) is the study of methods that integrate human intelligence and algorithmic search to help people better search, explore, learn from, and leverage information (Marchionini, 2006b). It comprises a number of sub-disciplines including exploratory search (White & Roth, 2009) which combines querying and browsing strategies to foster learning and investigation; information retrieval in context (Ingwersen & Järvelin, 2005) which considers aspects of the user or environment that are typically not reflected in query statements; and interactive information retrieval (IIR), which is primarily focused on the interactive communication processes between the main actors in retrieval operations: users, systems, and optionally, intermediaries (Ingwersen, 1992). All of these sub-disciplines share a common goal: understanding and improving the way that people leverage technology to derive value from information. HCIR affords searchers, their self-actualization, and their deep involvement in the search process, predominant roles in search system design, use, and evaluation.

Over the past few decades, search behavior has converged on a small number of tactics that transform the user into a passive information receiver rather than an active information seeker (Allan et al., 2012). HCIR systems empower searchers to be more proactive and critical thinkers during the information search process. To achieve this they must help users acquire better information search skills (through tutorials, tips, etc. (Moraveji et al., 2011; Bateman et al., 2012)) and provide better system support for information interaction and understanding. Marchionini (2006b) proposed the following design goals for search systems where users have more control in determining relevant results: (i) help people making sense of information (cf. Dervin, 1983); (ii) amplify and reward good intellectual effort; (iii) have flexible architectures; (iv) situate themselves within a broader information ecology of memories and tools; (v) support the information life-cycle from creation to use; (vi) support tuning by users, and (vii) should be engaging and enjoyable to use. By realizing many of these goals, HCIR systems can help people become more effective information seekers and consumers, and more informed individuals generally.

Existing capabilities of search systems already provide some support for the objectives listed above. Beyond presenting a ranked list of search results chosen with respect to the provided query, features such as spelling suggestions and suggested queries provide mechanisms to lead the user to potentially relevant content. Importantly from an HCIR perspective, control over selection and interpretation of the suggestions offered still remains with the searcher. Systems are also becoming increasingly aware of their users and their situations that extend beyond the query statements they issue. Personalization employs user models to individualize search results to a particular user's interests (Teevan et al., 2005). Relevance feedback (including implicit feedback from interaction behavior and biometrics from physiological sensors) enables search systems to develop richer representations of users' tasks and search interests by capturing relevance judgments explicitly from searchers (Rocchio, 1971) or mining them from their search interactions (Kelly & Teevan, 2003) or physiological signals (Feild et al., 2010). The search behavior of other users (primarily queries and clickthrough) can also be mined in the aggregate and used to help rank results (Joachims, 2002; Agichtein et al., 2006) or suggest paths to follow through document collections (Wexelblat & Maes, 1999; White et al. 2007). Recent research on mining finer-grained search interactions, such as mouse movements to estimate searcher's gaze attention, is a promising direction for richer behavioral modeling on the Web (Guo and Agichtein, 2010; Huang et al., 2011).

The application of information to garner insights and make decisions is another important requirement of HCIR systems. Information visualization, visual analytics, summarization, and collaboration methods help users make meaning from the results returned for a query. Beyond better ranking and query formulation support, search systems can also provide task-appropriate forms of data display (Shneiderman et al., 1997) that allow users to explore, analyze and synthesize information, and interact and engage with information in more meaningful ways, amplifying cognitive capabilities (Card et al., 1999). Summarization helps aggregate or compress search results into a more human-consumable form. Faceted search (Tunkelang, 2010) is one such form of summarization that enables users to navigate information hierarchically, transitioning from a category to its sub-categories, but choosing the order in which the categories are examined. Faceted navigation (Hearst, 2009), like taxonomic navigation, guides users by showing them available categories (or facets), but does not require them to browse through hierarchies that may not reflect their search intentions or map well to how they want to explore the data. Another summarization method is search-result clustering (Salton 1971), which groups similar or co-occurring documents or terms allowing for analysis of the clusters themselves (which can be informative in their own right) or use the clusters as a way to refine the search space. The output of summarization or analytics may be displayed as tables, charts, or summaries of aggregated data such as TreeMaps (Shneiderman, 1992) which may be updated dynamically based on users' manipulation of provided control mechanisms (Ahlberg & Shneiderman, 1994). Searchers may also seek answers from other searchers either as part of a collaborative search task (Morris & Horvitz, 2007) or in question-and-answer (Q&A) scenarios where questions are posed broadly to online communities such as Yahoo! Answers (answers.yahoo.com) or to a group of knowledgeable (expert) answerers via a more focused medium such as instant messaging (Horowitz & Kamvar, 2010).

In addition to providing enhanced system support, HCIR systems need to be evaluated in a way that reflects the requirements of their users. Most modern IR systems employ a ranked retrieval model, in which the documents are scored based on estimates of document relevance to a query. These systems are typically evaluated based on their performance over a set of benchmark queries in settings such as the Text Retrieval Conference (TREC) (Harman, 1993). Because of its emphasis on the role of human intelligence in the search process, HCIR requires different evaluation models. Early work on interactive IR leverages standard measures such as precision and recall but applies them to the results of multiple iterations of user interaction, rather than to a single query response. One example of this is the work of Koenemann and Belkin (1996), where they examine different levels of transparency in query suggestion. Other HCIR research, such as Borlund's IIR evaluation model, applies a methodology more reminiscent of HCI, focusing on key issues such as the characteristics of users and the details of experimental design (Borlund, 2001). More recent research on the evaluation of HCIR systems has shifted the focus toward the measurement of non-traditional, but nonetheless critical, aspects of the search process such as user engagement (Obrien & Toms, 2009), search satisfaction (Fox et al., 2005), time-based gain (Smucker and Clarke, 2012), learning as a function of exploration time (Rao & Card, 1994), or the rate of grasping variable properties or inter-variable relationships during exploratory data analysis (Pirolli & Rao, 1996). Establishing a set of trusted metrics is critical for the development of effective HCIR systems and we expect to see more of these metrics emerge over time – including some that are outlined in this special topic issue.

This special issue makes progress towards more sophisticated modeling of searcher needs, discusses the provision of support to help users, and covers issues in the design and evaluation of HCIR systems. It builds on the increased interest in HCIR, especially in light of the popular HCIR workshop series that has run annually for the past six years in the United States and has recently spawned a European counterpart in the EuroHCIR event. The goals of the special issue are two-fold: (i) foster the design of search systems that support complex problem solving, increase user control and responsibility, facilitate intellectual development, support the full information life cycle, and are engaging and enjoyable to use, and (ii) provide an outlet for significant previously-unpublished research on HCIR. The special issue articles describe complete pieces of work that aim to enhance the community's knowledge of existing search behaviors, demonstrate through sound experimentation how new interactive tools can assist searchers, or propose and validate new evaluation metrics and methods for HCIR scenarios. The collection of articles in this special topic issue covers a range of pertinent topics including exploration, engagement, task effects, inferring knowledge from behavior, information visualization and collaborative search.

The acquisition of information and the search interaction process is influenced strongly by a person's use of their knowledge of the domain and the task. Two of the accepted articles in this special topic issue specifically focused on searcher knowledge. In *Examining Users' Knowledge Change in the Task Completion Process* Jingjing Liu and colleagues describe a study of changes in searchers' topic knowledge during the completion of search tasks. They studied knowledge changes over time (across three search sessions) and demonstrated that users' knowledge generally increases after each search session. They also found that knowledge was correlated with participants' perceptions of task attributes and accomplishment, and that task type affects several aspects of knowledge levels and knowledge change. Also addressing the challenge of modeling user knowledge, *Inferring User Knowledge Level from Eye Movement Patterns* by Michael Cole and colleagues, presents an approach to infer domain knowledge by modeling a user's information acquisition process during search using only measurements of eye movement patterns. They performed a user study in the genomics domain where they compared participants' self-ratings of domain knowledge with measures of cognitive effort associated with reading patterns during search tasks. The results show correlations between the cognitive effort due to reading and an individual's level of domain knowledge. The authors constructed exploratory regression models capable of predicting knowledge levels based on real-time measurements of eye movement patterns during a task session.

Accurately measuring the effect of search systems on user perceptions is essential for improving existing HCIR systems and developing new systems. In *Measuring Engagement in Search Systems using the User Engagement Scale*, Helen O'Brien and Elaine Toms present methods to measure the user experience during search, specifically focusing on user engagement. They propose an experimental instrument, the user engagement scale (UES), that measures the user experience in a more comprehensive way than only satisfaction, allowing for the capture of dynamic, multifaceted experiences that evoke pragmatic and hedonic needs, expectations, and outcomes associated with exploratory search. The UES captures six factors, including sub-scales pertaining to usability, aesthetics, and attention. To perform rigorous reliability and validity testing of the UES, the authors ran a laboratory study with a large number of participants performing complex search tasks on a new search interface. They present findings from their analysis of the scale's performance and showed that the some of the sub-scales were stable and others could be merged to form a single factor. In *A Cross-Domain Analysis of Task and Genre Effects on Perceptions of Usefulness*, Luanne Freund presents the results of two studies of an approach to identify more useful documents through the relationships between searchers' tasks and document genres. This research makes progress towards developing search systems that can distinguish between information that is on topic and information that is useful, i.e., suitable and applicable to the tasks at hand. Freund describes a questionnaire and a user study conducted in two domains that provide evidence that perceptions of usefulness are dependent upon information task type, document genre, and the relationship between these two factors. Aligned with the studies on user knowledge described earlier, Freund also shows that people's domain knowledge affects their perceptions of usefulness.

As well as measuring aspects of the search process, HCIR is very much focused on the development of new systems to help people search and learn more effectively. In *CIDER: Concept-based Image Diversification, Exploration, and Retrieval*, Orland Hoeber and colleagues describe support for image search that first performs automatic query expansion using Wikipedia as the source knowledge base, and then organizes the diverse search results using both the conceptual information associated with each image, and the visual features extracted from the images. This, coupled with a hierarchical organization of the concepts, provides an interactive interface that takes advantage of the searchers' abilities to recognize relevant concepts, filter and focus the search results based on these concepts, and visually identify relevant images while navigating within the image space. The authors outline the key features of their system, and present the results of a preliminary user study that demonstrates its potential benefits to searchers conducting image retrieval tasks. In *Adaptive Visualization for Exploratory Information Retrieval*, Jae-wook Ahn and Peter Brusilovsky propose the incorporation of interactive visualization into personalized search and introduce an adaptive-visualization-based search system. The system aims to enhance the effectiveness of personalization by allowing users to interact with the system and learn more about the problem at hand. The authors evaluate their system in a user study and show that it can improve the precision and the productivity of personalized search while helping users discover more diverse sets of information.

Moving beyond the individual searcher, collaboration is an important aspect of human-centered IR. In *Let's Search Together, But Not Too Close! An Analysis of Communication and Performance in Collaborative Information Seeking*, Roberto González-Ibáñez and colleagues investigate different types of communication channels used by information seekers on a collaborative project. They described a user study with 30 pairs (60 participants) in three different conditions: co-located, remotely located with text chat, and remotely located with text as well as audio chat, in an exploratory search task. They show that pairs of collaborators who are remotely-located were more effective at exploring diverse information, and that adding audio support for remote collaboration helped participants to lower both their cognitive load as well as negative emotions compared to those working in the same space. The authors also show how these findings could help design more effective systems for collaborative information seeking tasks using adequate and appropriate communication.

Moving forward, the models of behavior that the HCIR community develops and the systems that they build based on them should keep pace with technological advances and user demands. Search is expanding into new domains (e.g., leisure, diagnosis, and social) and how people are interacting with information is fundamentally changing (Hearst, 2011). New modalities (e.g., natural user interaction such as multi-touch on tablets and smartphones, gestural interactions, speech input and output, augmented reality), different form factors (e.g., smartphones, laptops, tablets, immersive search and exploration environments), and are handling a broader range of tasks than ever before. Big data is also emerging as a critical component of current and future search systems, both in terms of searching new data sources (e.g., statistics, multimedia, sensor streams) and leveraging new types of data (e.g., search logs and social network graphs) to improve search performance and offer recommendations based on the aggregated behavior of other users. Earlier we pointed to some ways in which data can be helpful in directing users toward potentially-relevant results, but there is still significant opportunity for searchers to learn from the activities of the masses or particular subsets with shared interests or more domain knowledge (experts).

The coming years will bring a revolution in the design of search technology driven by new capabilities (such as those highlighted above) but also a shift in people's expectations for the support that search technology should provide to them. Beyond simply finding information, more focus will be placed on enhanced capabilities to empower people to search more effectively, efficiently solve complete search tasks (not just single queries), and facilitate their cognitive development. We envisage that HCIR research, such as that presented in this special issue, will be central to this change and will play a pivotal role in creating the enlightened society that could follow.

References

Agichtein, E., Brill, E., and Dumais, S.T. (2006). Improving web search ranking by incorporating user behavior information. *Proc. SIGIR*, 19–26.

Ahlberg, C. and Shneiderman, B. (1994). Visual information seeking: tight coupling of dynamic query filters with starfield displays. *Proc. SIGCHI*, 313–317.

Allan, J. et al. (2012). Frontiers, challenges, and opportunities for information retrieval: report from SWIRL 2012 strategic workshop on information retrieval. *SIGIR Forum*, 46(1): 2–32.

Bateman, S., Teevan, J., and White, R.W. (2012). The search dashboard: how reflection and comparison impact search behavior. *Proc. SIGCHI*, 1785–1794.

Borlund, P. (2003). The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. Information Research, 8(3): 152.

Card, S., Mackinlay, J.D. and Shneiderman, B. (1999). *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann: San Francisco, CA.

Dervin, B. (1998). Sense-making theory and practice: an overview of user interests in knowledge seeking and use. *Journal of Knowledge Management*, 2(2): 36–46.

Engelbart, D. (1962). *Augmenting Human Intellect: A Conceptual Framework*. Summary Report AFOSR-3233. Menlo Park, CA: Stanford Research Institute.

Feild, H., Allan, J. and Jones, R. (2010). Predicting searcher frustration. *Proc. SIGIR*, 34–41.

Fox, S., Karnawat, K., Mydland, M., Dumais, S.T., and White, T. (2005). Evaluating implicit measures to improve web search. *ACM TOIS*, 23(2): 147–168.

Guo, Q. and Agichtein, E. (2010). Ready to buy or just browsing? Detecting Web searcher goals from interaction data. *Proc. SIGIR*, 130–137.

Harman, D. (1993). Overview of the first TREC conference. *Proc. SIGIR*, 36–47.

Hearst, M. (2009). *Search User Interfaces*. New York, NY: Cambridge University Press.

Hearst, M. (2011). ‘Natural’ search user interfaces. *CACM*, 54(11): 60–67.

Horowitz, D. and Kamvar, S.D. (2010). The anatomy of a large-scale social search engine. *Proc. WWW*, 431–440.

Huang, J., White, R.W., and Dumais, S.T. (2011). No clicks, no problem: using cursor movements to understand and improve search. *Proc. SIGCHI*, 1225–1234.

Ingwersen, P. (1992). *Information Retrieval Interaction*. London, U.K.: Taylor Graham.

Ingwersen, P. and Järvelin, K. (2005). *The Turn: Integration of Information Seeking and Retrieval in Context*. Secaucus, NJ: Springer-Verlag.

Joachims, T. (2002). Optimizing search engines using clickthrough data. *Proc. SIGKDD*, 133–142.

Kelly, D. and Teevan, J. (2003). Implicit feedback for inferring user preference: a bibliography. *ACM SIGIR Forum*, 37(2): 18–28.

Koenemann, J. and Belkin, N.J. (1996). A case for interaction: a study of interactive information retrieval behavior and effectiveness. *Proc. SIGCHI*, 205–212.

Marchionini, G. (2006a). Exploratory search: from finding to understanding. *CACM*, 49(4): 41–46.

Marchionini, G. (2006b). Toward human-computer information retrieval. *Bulletin of the American Society for Information Science and Technology*, June/July.

Moraveji, N., Russell, D.M., and Mease, D. (2011). Measuring improvement in user search performance resulting from optimal search tips. *Proc. SIGIR*, 355–364.

Morris, M.R. and Horvitz, E. (2007). SearchTogether: an interface for collaborative web search. *Proc. UIST*, 3–12.

O'Brien, H.L. and Toms, E. (2010). The development and evaluation of a survey to measure user engagement. *JASIST*, 61(1): 50–69.

Pirolli, P. and Rao, R. (1996). Table lens as a tool for making sense of data. *Proc. Working Conference on Advanced Visual Interfaces*, 67–80.

Rao, R. and Card, S.K. (1994). The table lens: merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. *Proc. SIGCHI*, 318–322.

Rocchio, J.J. (1971). Relevance feedback in information retrieval. In *The SMART Retrieval System - Experiments in Automatic Document Processing*, 313–323.

Salton, G. (1971). Cluster search strategies and the optimization of retrieval effectiveness. In *The SMART Retrieval System - Experiments in Automatic Document Processing*, 223–242.

Shneiderman, B. (1992). Tree visualization with treemaps: a 2-d space-filling approach. *ACM Transactions on Graphics*, 11(1): 92–99.

Shneiderman, B., Byrd, D. and Croft, W.B. (1997). Clarifying search: a user-interface framework for text searches. *D-Lib Magazine*.

Smucker, M. and Clarke, C.L.A. (2012). Time-based calibration of effectiveness measures. *Proc. SIGIR*, 95–104.

Teevan, J., Dumais, S. and Horvitz, E. (2005). Personalizing search via automated analysis of interests and activities. *Proc. SIGIR*, 449–456.

Tunkelang, D. (2009). *Faceted Search*. Morgan Claypool.

Wexelblat, A. and Maes, P. (1999). Footprints: history-rich tools for information foraging. *Proc. SIGCHI*, 270–277.

White, R.W., Bilenko, M., and Cucerzan, S. (2007). Studying the use of popular destinations to enhance web search interaction. *Proc. SIGIR*, 159-166.

White, R.W. and Roth, R.A. (2009). *Exploratory Search: Beyond the Query-Response Paradigm*. Morgan Claypool.