HCIR 2009: The Third International Workshop on Human-Computer Interaction and Information Retrieval

Bill Kules
The Catholic University of America
kules@cua.edu

Daniel Tunkelang
Endeca
dt@endeca.com

Ryen W. White
Microsoft Research
ryenw@microsoft.com

Abstract

This report describes HCIR 2009, the third international workshop on Human-Computer Interaction and Information Retrieval (HCIR), held in October 2009 at The Catholic University of America in Washington, DC. The workshop attracted over 50 participants from across the world and was the largest HCIR workshop to date. The event brought together representatives from academia, industry, and government to discuss research, present work in process, and advance ideas in the area of HCIR. The workshop consisted of a keynote presentation, panel presentations, a poster session, and two guided discussion sessions.

1 Introduction

Gary Marchionini coined the term "human–computer information retrieval" (HCIR) in a 2005 lecture, advocating an approach that "aims to empower people to explore large-scale information bases but demands that people also take responsibility for this control by expending cognitive and physical energy" [8]. He was not the first person to try to unify the areas of information retrieval (IR) and human-computer interaction (HCI): in 1996 and 1998, a pair of workshops at the University of Glasgow sought to address the overlap between these two fields [6]. More recently, Salton Award winner Sue Dumais was specifically recognized for three decades of "leadership in bridging the fields of information retrieval and human computer interaction" [3].

Nonetheless, in 2007, a number of us who work as researchers and practitioners in the overlap of these two fields bemoaned the lack of a community and scholarly venue for advancing it. We were (and still are) regular attendees at established conferences like SIGIR and SIGCHI, as well as smaller conferences like JCDL and UIST, but those conferences often confine our primary interests to a single session or workshop.

This need birthed the first HCIR workshop, held in October 2007 at MIT in Cambridge, MA, co-sponsored by Endeca and IBM Research. The workshop created sufficient interest to inspire a second workshop at Microsoft Research in Redmond, WA, and now a third at The Catholic University of America in Washington, DC. The authors of this report (Bill Kules, Daniel Tunkelang, and Ryen White) serve as a steering committee for the annual workshops.

This year's workshop is especially timely. Mainstream web search engines are taking an increasing interest in user interaction, while explicit support for interactive IR—particularly faceted search—

have become ubiquitous in site and enterprise search settings. Nonetheless, there is a dire need for models, tools, and evaluation methods to support such innovative efforts.

2 The Workshop

The workshop was a highly interactive event that included a variety of different activities, including a keynote, poster boasters, interactive poster sessions, system demonstrations, a panel, and two guided discussion sessions. The proceedings are available on the workshop website: http://cuaslis.org/hcir2009/.

2.1 Keynote

Ben Shneiderman, professor at the University of Maryland and founding director of the Human-Computer Interaction Laboratory, delivered an excellent keynote on "The Future of Information Discovery." He began by asserting that, while simple fact-finding is largely addressed by today's search engines, we need better tools to address extended fact-finding tasks (with vague queries), tasks involving exploration of availability (with vague result requests), open-ended browsing and problem analysis (with hidden assumptions), and mismatches between information needs and available metadata (requiring exhaustive search). Shneiderman suggested that HCIR systems should enrich query formulation, expand result management, enable long-term effort, and enhance collaboration—while enabling users to deal with concerns such as completeness (especially in legal and medical domains), absence (proving non-existence), outliers (making unexpected connections), and bridging (connecting disciplines). He urged HCIR researchers to embrace a broader vision, e.g., relating our work to the UN Millennium Development Goals. He presented examples of HCIL work to support medical research and other areas of broad social concern. Finally, he urged the HCIR community to focus on multi-dimensional in-depth long-term case studies (MILCs) in addition to controlled experiments (cf. Science 2.0 [10]).

2.2 Panel

The panel session consisted of five oral presentations:

- *Usefulness as the Criterion for Evaluation of Interactive Information Retrieval*
  M. Cole, J. Liu, N. J. Belkin, R. Bierig, J. Gwizdka, C. Liu, J. Zhang, and X. Zhang
- *Modeling Searcher Frustration*
  H. Feild and J. Allan
- *Query Suggestions as Idea Tactics for Information Search*
  D. Kelly
- *I Come Not to Bury Cranfield, but to Praise It*
  E. Voorhees
- *Search Tasks and Their Role in Studies of Search Behaviors*
  B. M. Wildemuth and L. Freund

Jingjing Liu argued that IR system evaluation should consider both task success and the value of support given over the entire information seeking episode, and suggest that relevance-based measurements fail to address these requirements. Henry Feild described work on modeling and detecting user frustration during search based on inputs from physical sensors. Diane Kelly presented the results from two studies that examined people's use of query suggestions while searching for open-ended search topics and how usage varied according to topic difficulty. Ellen Voorhees argued that Cranfield-style [4] experimentation is critical to the study of interactive (user-in-the-loop) retrieval in order to manage trade-offs among realism, experimental power, and cost. Finally, Barbara Wildemuth reported on an ongoing analysis of search tasks that have been used in experimental

search studies and reviewed a number of typologies of search tasks currently in use to help guide the design of search tasks for use in future studies.

A question-answering session followed the panel presentations. Issues covered included: the need to study traces of user interaction rather than just queries, documents, and relevance judgments; the value of social information; frustration analysis, including the need to consider users' emotional priors when conducting frustration analysis; simulations of user as a way to complement the Cranfield methodology; and mixed methods and multidimensional analyses as a way to more completely evaluate HCIR systems.

2.3 Posters and Demonstrations

Twenty-five attendees presented posters at the workshop. The workshop began with a "poster boaster" session that gave presenters an opportunity to pitch their work to all attendees in a one-minute presentation. During the one-hour poster session, participants mingled, browsed posters, interacted one-on-one with poster presenters, and watched system demonstrations.

2.4 Group Discussion

Two workshop wide discussion sessions closed out the workshop. In the first session, Ryen White led a discussion on the theory of HCIR, including models and evaluation. In the second session, Bill Kules led a discussion on the practice of HCIR, including the design of tools.

2.4.1 Theory, Models, and Evaluation

This session covered numerous topics, including affect, models, negative search, and evaluation.

Attendees agreed that affect is a fertile but challenging research area and cited Carol Kuhlthau's research on the role of emotion in the information seeking process [7], as well as discussing the capture of affect through sensors, facial expressions (e.g., [9]), or explicitly asking users.

Attendees also discussed and the role of modeling in HCIR—particularly, the need to consider the user and the user experience in any models of the search process (cf. Bates's work on the boundary between users and systems [1]). The explicit support of user actions beyond the retrieval of search results is a key differentiator between HCIR and IR. There is also the need to move beyond purely cognitive models, and consider perspectives such as "embodied minds" (i.e., that we are mental beings but we have affective sides, we have bodies and we are embedded in environments) and query models (i.e., using motivation plus topic in system development and evaluation rather than topic alone). The need for a standard database of user interactions, similar to the Lemur Query Log Toolbar was a widely held concern.

Attendees also discussed the problem of negative search (determining the absence of information). Negative search is particularly important in legal, patent, and medical domains and has been understudied in the IR community, with the exception of the TREC Question-Answering Track, which investigated factoid absence. However, real world problems are much richer than proving that a factoid is absent from a document collection.

Attendees agree that HCIR research needs to develop more robust evaluation measures. The Cranfield model (and associated measures such as precision and recall), task completion time, and GOMS (Goals, Operators, Methods, and Selection) [2] were all discussed as possible ways to evaluate in HCIR. Attendees also discussed reductionist approaches to interaction, such as looking in the middle of tasks and measuring interim aspects with thinner "slices." Log analysis was suggested as a way to thin slice and find successes and dead ends, but for tractability, evaluation may need to be tied to goal of the session rather than its subcomponents. Rich user models that associate cognitive

processes with resource selection among other factors could facilitate the development of more refined measurement processes.

Attendees also discussed how the HCIR community should encourage the use of metrics beyond precision and recall (e.g., novelty, diversity, utility). Doing so requires an attitude shift from both authors and reviewers, Reviewers should encourage authors to motivate (and validate) the use of particular metrics, explain in detail why they selected a metric, and prove causal relationships between findings and independent variable(s). Meanwhile, referees should be more open to papers that use non-standard (but appropriate) metrics and to papers with negative results.

2.4.2 Tools and Practice

This session focused on the more practical aspects of HCIR research and development.

Attendees agreed that HCIR tools are not just about powerful search interfaces: indeed, the increasing adoption of faceted search in commercial applications demonstrates the value of relatively simple tools to support interaction. Moreover, search engines are starting to compete on interface innovation, rather than simply on ranking algorithms.

Attendees also acknowledged the elephant in the room: when most people think of search, they think in terms of popular web search engines. The HCIR community should be at the forefront of creating, evaluating, and publicizing tools and techniques that help ordinary people be more successful at satisfying their information needs.

Other topics of discussion included where and how HCIR relates to question answering (e.g., are they antitheses or complementary topics?), whether HCIR researchers should investigate spam detection/suppression, and how HCIR researchers can make use of standard psychometric measures. There are also pragmatic concerns facing HCIR researchers: the information most useful to advancing the field also presents challenges to user privacy. As tools become more powerful and retain more information about information seekers, the question of protecting that information and making it appropriately available (to people or systems) becomes more urgent.

The discussion concluded with the broader question of how to define HCIR, a topic that surfaced multiple times during the day. Do we define it by the key problems we address or by a set of techniques? What are the boundaries? HCIR studies actions and interactions with information: how people find and use information when mediated by technology. Search is just one part of the information seeking episode. HCIR includes collaboration, social interactions, and personal information management (PIM). One proposal was to embrace the three-level model of Ingwersen and Järvelin [5]. Whether the scope extends to higher level work tasks or is more narrowly limited to the information seeking tasks remains an open question. Attendees agreed that it is important to establish a clear sense of identity while embracing the diversity of our research.

3 Concluding Remarks

This was the third HCIR workshop, and as a community we are developing a shared understanding of each other's work. We see the rich interconnections between information needs, information seeking behaviors, exploratory search, collaborative search, recommender systems, as well as visualization tools, user interface design principles and information organization. We need to continue developing our understanding of these areas to develop a strong core for what HCIR is. At the same time, we need to keep an eye looking outward – to see how our work interacts with and contributes to other fields. Our work is inherently interdisciplinary and we need to nurture that. We need to think creatively and collaboratively, to make those connections, and share our ideas for where the HCIR community is going.

It is essential that we develop an agenda of tasks and types of tasks that HCIR investigates and supports. Doing so will yield practical benefits for researchers who develop and validate tasks for evaluation. Moreover, some degree of standardization will make evaluations more comparable across systems and studies. Finally a more formal agenda will help clarify what problems we are focusing on, which in turn will establish how HCIR differs from and contributes to other areas of research.

The next HCIR Workshop will be co-located with the 2010 Information Interaction in Context Symposium (IIiX 2010), on Sunday, August 22, 2010, in New Brunswick, NJ, USA. We hope to see you there!

References

[1] Bates, M.J. (1990). Where should the person stop and the information search interface start? *Information Processing and Management*, 25 (5), 575-591.

[2] Card, S., Moran, T.P. and Newell, A. (1983). *The Psychology of Human Computer Interaction*. Lawrence Erlbaum Associates.

[3] Citation from 2009 Gerard Salton Award presentation (www.sigir.org/awards/awards.html).

[4] Cleverdon, C.W., Mills, J. and Keen, E.M. (1968). *Factors Determining the Performance of Indexing Systems*. Two volumes, Cranfield, England.

[5] Ingwersen, P., & Järvelin, K. (2005). *The Turn: Integration of Information Seeking and Retrieval in Context.* Springer, Dordrecht.

[6] Johnson, C. and Dunlop, M., Eds. (1998). Special issue on HCI and information retrieval. *Interacting with Computers*, 10 (3).

[7] Kuhlthau, C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42 (5), 361-371.

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

[9] Nahl, D. and Bilal, D., Eds. (2007). *Information and Emotion: The Emergent Affective Paradigm in Information Behavior Research and Theory.* Medford, NJ: Information Today.

[10] Shneiderman, B. (2008). Science 2.0. *Science*, 319, 1349-1350.