
WORKSHOP REPORT

HCIR 2011: The Fifth International Workshop on Human-Computer Interaction and Information Retrieval

Robert Capra¹
University of North Carolina
at Chapel Hill
rcapra@unc.edu

Gene Golovchinsky
FX Palo Alto
Laboratory
gene@fxpal.com

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

Dan Russell
Google
drussell@google.com

Catherine L. Smith
Kent State University
csmit141@kent.edu

Daniel Tunkelang
LinkedIn
dtunkelang@linkedin.com

Ryen W. White
Microsoft Research
ryenw@microsoft.com

¹ Workshop organizers are listed in alphabetical order.

For more information about the workshop series, please visit <http://hcir.info>.

Submission details for the special issue of Information Processing and Management on HCIR are available at <http://hcir.info/hcir-2011/ipm-special-issue>. Submission deadline: January 6, 2012.

Abstract

This report describes the 2011 Workshop on Human-Computer Interaction and Information Retrieval. Now in its fifth year, the workshop was held in October, in Mountain View, CA. The event brings together researchers from academia, industry, and government and a range of disciplines for in-depth discussions in an informal atmosphere. The workshop continues to grow, with around 100 attendees this year. We continued the HCIR Challenge, this time focusing on the problem of information availability, with four in-depth system demonstrations, and audience selection of a challenge winner.

1 Introduction

Human-computer information retrieval (HCIR) is the study of information retrieval (IR) techniques that integrate human intelligence and algorithmic search to help people explore, understand and use large information spaces better. The fields of human-computer interaction (HCI) and IR have both developed innovative techniques to address the challenge of navigating complex information spaces, but their insights frequently fail to cross disciplinary boundaries. Human-computer information retrieval has emerged in academic research and industrial practice to bring together research in the fields of IR and HCI, in order to create new kinds of search systems that depend on continuous human control of the search process (Marchionini, 2006).

The HCIR workshop provides a venue for in-depth discussion of models, tools, and evaluation methods at the intersection of human factors and search. The event has run annually since 2007 and has attracted growing interest from HCI and IR researchers, serving as bridge between the two communities. Its success inspired the first EuroHCIR workshop, held in July 2011. The fifth HCIR workshop (HCIR 2011) was held on Thursday, October 20, 2011 at Google headquarters in Mountain View, CA. This year's workshop attracted around 100 attendees, and featured presentations and posters, with lots of opportunity for discussion among attendees.

2 The Workshop

Activities at the workshop included a keynote, paper presentations (with a mix of long and short talks), a highly-interactive poster session, and the HCIR Challenge in which participating groups built systems to help address the information availability problem using the CiteSeer data.

2.1 Keynote

Gary Marchionini, who coined the term “HCIR,” gave the keynote address “HCIR: Now the Tricky Part.” Gary is the Dean and Cary C. Boshamer Professor at the School of Information and Library Science, University of North Carolina at Chapel Hill. The talk used three case studies of HCIR projects to focus the audience’s attention on one of the main challenges of HCIR: how to evaluate the systems we build. A case study of *Open Video* (open-video.org) introduced the notion of surrogates, which are representations of metadata designed for human consumption. One significant challenge, particularly for video objects, is how to evaluate surrogate effectiveness. While most evaluations are laboratory studies, there are questions around whether these findings retain validity in naturalistic settings. The second case study focused on the *Relation Browser*, a faceted browsing tool. The Relation Browser is designed to reveal relationships across facets, and supports exploration through browsing rather than keyword search. The system has undergone many revisions and redesigns over a decade. Gary described how his (and his laboratory’s) experience with the project raised several issues that are applicable much more broadly, e.g., How does one capture the design rationale that goes into evolving the system over time? How to show the benefits of HCIR interfaces when users prefer interfaces with which they are already familiar? The third case study, based on the ongoing ResultsSpace project, explored whether we can better represent search results than by using lists of snippets, and touched on issues of awareness in collaborative search.

Having presented the case studies, Gary raised some broader issues for the HCIR field. While he admitted that most of his work focused on laboratory studies, he recognized the need for field observations. In fact, he suggested that neither approach is truly useful in isolation. A process that starts with qualitative observations to generate hypotheses that are then tested in the laboratory is one effective way to integrate the methods, but Gary also suggested proceeding in the opposite direction.

Gary also identified the following challenges:

1. Query quality: Can we assess the quality of a query, which is often the first signal of user behavior? Can this be done pre-retrieval as well as post-retrieval? How can we elicit human judgments of query quality?
 2. Behavior as evidence: What are the various ways in which searchers’ behavior can be converted into implicit relevance feedback? Can we match behaviors to queries to infer what people might be interested in?
 3. How do we design and evaluate surrogates that help people perceive, recognize, gist, understand, interpret, analyze and evaluate search results?
 4. How do the tools we build affect searchers’ cognitive load? As we add more tools, are we helping or hurting? Does collaboration simply increase the load further without offering tangible benefits, or is there a useful tradeoff?
 5. How do we measure session quality and search quality? Recall and precision are point-based measures that do not capture the process characteristic of session-based search. How do we adapt them to the session as a unit of interaction (as opposed to a query)? What other metrics are appropriate for session-based measurement?
-

-
6. Finally, he touched on the tricky issue of recording and capturing experimental traces: annotated logs, video recordings, data sets, etc. Should researchers publish traces of system use to characterize participants' search activities?

The talk also included an interactive interlude, where Gary showed the audience a graph with two lines. The vertical axis represented an undefined performance metric (although options included precision, recall, and even the number of user smiles – a reasonable proxy for satisfaction), and the horizontal axis was time. One of the lines showed steady progress from lower-left to upper-right, while the other took a more erratic route between the same starting and ending point. Gary's question to the audience was which line represented better, or perhaps more desirable, system performance. A number of suggestions were offered, including the benefits of steady progress, of relieving frustration through surprise (the "aha!" moment reflected in a sharp upward spike of the erratic line), of achieving maximum performance if only for a short time, etc. There was some consensus that the erratic line reflected a learning process that is often characteristic of exploratory search.

The keynote was a great summary of some accomplishments of HCIR research and some challenges ahead. We expect that next year's presentations will offer some insights into these thorny issues!

2.2 Papers and Posters

We accepted submissions for presentation as talks (either 15 minutes or 5 minutes in duration), or as poster presentations, which were presented during the dedicated poster session.

2.2.1 Papers

We accepted fourteen papers for oral presentation. Six of the accepted papers were presented as 15-minute talks and eight were presented as 5-minute talks. We divided the talks into two sessions.

At the first session, there were three 15-minute presentations followed by four 5-minute "lightning" presentations. To start, Sofia Athenikos from Collective[i] presented details of a search system called *PanAnthropon FilmWorld* that supports semantics-based retrieval on entities, facts, and relations. In a user study with movie-related query tasks, she showed that the system outperformed the Internet Movie Database (IMDB) in terms of precision/recall and user ratings. Next, Chang Liu from Rutgers University described work that explored the effect of task difficulty and domain knowledge on dwell times. She found that dwell time on search engine result pages was a good indicator of task difficulty, and reported an interesting, but unexpected result that domain experts had shorter dwell time on content pages for difficult tasks than for easy ones. Then Jingjing Liu from Southern Connecticut State University presented research that explored how users gain knowledge in different types of multi-session search tasks. She reported that users gained knowledge across search sessions, and found differences between general tasks and sub-tasks in terms of patterns of knowledge gained.

Next were the 5-minute lightning talks. Mark Smucker from the University of Waterloo described research about how users examine and process ranked document lists, finding two main approaches: fast and liberal, and slow and neutral. Jinyoung Kim from the University of Massachusetts at Amherst presented research on an associative browsing model of personal information for a known-item finding task. He described results of an evaluation using a simulated user model that found that users' level of knowledge affected their finding behavior. Bill Kules from The Catholic University of America discussed results of a study that used a retrospective stimulated recall technique to gather participants' self-reported stages of search during exploratory searches using an online library catalog. Using a graphical visualization of the stages search, he illustrated patterns in the data. Michael Cole from Rutgers University presented research that explored the use of eye-tracking to measure domain knowledge in search. He reported results showing correlations between domain knowledge and eye movement patterns related to cognitive effort.

The second presentation session also contained papers on a broad range of topics. In the first of the 15-minute presentations in the session, Luanne Freund from the University of British Columbia presented work on the relationship between task type, document genre, and document usefulness. She described a user study in the e-government domain that showed that usefulness varied significantly by task type and by the genre of the search results; the genre effect was significant for tasks focused on learning, deciding, and doing, but not for problem-solving or fact-finding. Gene Golovchinsky from FX Palo Alto Laboratory described work on designing tools to support explicit, intentional collaboration in small teams. He focused on the interplay between the algorithmic mediation of collaboration and mediated communication among team members, and argued that the characteristics of a group's information need called for different design decisions. Alyona Medelyan from Pingar presented the findings of a user study where they asked bioscientists to evaluate the utility of autocomplete, query expansions, faceted refinement, related searches, and results previews in new pilot interfaces and publicly available search systems. She and co-author Anna Divoli found that different features were preferred for different tasks, and there was a strong preference for prototypes with just a few facets generated from the query or the matching documents.

Robert Capra from the University of North Carolina at Chapel Hill kicked off the set of four 5-minute lightning presentations describing the *Diamond Browser*, a system built to rapidly design, test, and conduct evaluations of faceted search interface designs for mobile devices. He described the system architecture and two initial interfaces that he, and co-author Jason Raitz, had developed. Keith Bagley from IBM and the University of Massachusetts at Lowell described a framework and prototypical system called *Conceptual Mile Markers* that aims to reduce the "time to value" for a search by matching a user query with existing search paths, thereby allowing users to profit from the "wisdom of the masses" and reuse previously-successful search activity. Additionally, he suggested that search trails may be saved, shared, and retrieved later as a style of lightweight personalization without the need for explicit user tracking. Xiaojun Yuan from the State University of New York at Albany presented an investigation of the effect of cognitive styles on users' search task performance using an information system called *Web of Science*. Her results demonstrated that users' cognitive styles did not impact their search performance. Michael Zarro from Drexel University presented a search system enhanced with social tags and Medical Subject Heading (MeSH) terms, and reported the results of a user experiment exploring its use. The prototype interface exposed terms to users that could be used to modify search queries, and was rated useful by participants in the user study.

2.2.2 Posters

Twenty-eight papers were selected for poster presentation, including position papers, reports on work in progress, and research results. Presenters introduced their work in one-minute poster-boasters at the beginning of the workshop, and lively discussions followed in a one-hour poster session later in the day. The posters covered a broad range of applications, including digital humanities, enterprise search, and image, music and database retrieval. Four posters focused on social navigation, collective search, and collaborative search, including one poster on schoolchildren's collaborative search. Work on a variety of interface types was presented, including conversational and touch interfaces, faceted designs, and interfaces for spoken-language search. Several posters presented work on information visualization in search interfaces, including a visualization method for the analysis of user behavior. Three posters addressed client-side data collection methods, including one proposal for a living laboratory for the collection of "task cases" for the evaluation of personal search systems. Other topics included an analysis of health websites, a survey on mobile phone use, discovery of topic trends using social tagging and citations, effects of cognitive capacity on search effort, and the effect of native language on search success.

2.3 The HCIR Challenge

Following its successful introduction at the 2010 workshop, the HCIR Challenge ran for a second year. This year it focused on the problem of information availability, where recall is paramount. Information availability is of concern when information seekers face uncertainty about whether the information of interest is available at all. Instances of this problem include some of the highest-value information tasks, such as those facing national security and legal/patent professionals, who may invest a large amount of time searching to determine whether the desired information exists. The corpus used for the challenge was the CiteSeer digital library of scientific literature, containing over 750,000 documents with rich meta-data on documents, authors, and citations (citeseer.ist.psu.edu). We are grateful to the CiteSeer team for providing participants with access to a snapshot of the data.

There were four challenge entries, each of which was presented as a talk at the workshop. Claudiu Firan from the University of Hanover demonstrated the *FreeSearch* system, which helps users proactively and unobtrusively by predicting their needs and providing precise suggestions, and requires no knowledge of the metadata schema. Gene Golovchinsky from the FX Palo Alto Laboratory described *Querium*, and showed how its interface and search engine can be used to search for documents in an open-ended, exploratory task via support such as relevance feedback, faceted search, query fusion, and the search history, as well as commenting and overview functions. Edmond Brian from Visual Purple showed how their *GisterPRO* system uses sophisticated visualization techniques to support the information seeking process. Ron Daniel, Matthew Corkum, and Keith Gutfreund from Elsevier Labs described the *Query Analytics Workbench*, a search application that incorporates both exploration and progressive refinement activities to guide the user through query construction.

There was intense competition among the entries. However, it came down to a close call between the Query Analytics Workbench and Querium. Despite the Elsevier team's impressive functionality and animated presentation, Querium's simpler interface and application of ranked fusion won the day in an audience vote. Our congratulations to Gene and Abdigani, this year's challenge winners!

3 Concluding Remarks

HCIR 2011 was an extraordinary success, attracting a record number of submissions (over 50) and a record number of attendees (around 100). The event gave researchers from academia and industry an opportunity to present their work and discuss ideas in an informal and interdisciplinary setting. The feedback we have received from attendees has been extremely positive, and we are grateful to everyone who contributed to the event's success. Our special thanks goes to Google for hosting the event, to our other sponsors, Microsoft Research, Endeca, and Kent State University, and in particular to all of the authors and presenters of papers and posters, who all did an outstanding job in both conducting such great research and in presenting it in an amazingly accessible way at the workshop.

We are already discussing HCIR 2012. We do not have many details as yet, but we are planning an event that is even bigger and better than the workshop this year. Look for more information about the venue, the format, paper submissions, etc. in the next few months. In the meantime, we look forward to receiving your articles for the special issue of Information Processing and Management on HCIR; January is just around the corner, so there is plenty to keep you all busy as our 2012 plans solidify! ☺

4 References

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