Full text available at: http://dx.doi.org/10.1561/1500000012

Methods for Evaluating Interactive Information Retrieval Systems with Users

Full text available at: http://dx.doi.org/10.1561/1500000012

Methods for Evaluating Interactive Information Retrieval Systems with Users

Diane Kelly

University of North Carolina at Chapel Hill
Chapel Hill, NC
USA
dianek@email.unc.edu

the essence of knowledge

Boston – Delft

Full text available at: http://dx.doi.org/10.1561/1500000012

Foundations and Trends® in Information Retrieval

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is D. Kelly, Methods for Evaluating Interactive Information Retrieval Systems with Users, Foundation and Trends® in Information Retrieval, vol 3, nos 1–2, pp 1–224, 2009

ISBN: 978-1-60198-224-7
© 2009 D. Kelly

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1-781-871-0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Full text available at: http://dx.doi.org/10.1561/1500000012

Foundations and Trends® in Information Retrieval
Volume 3 Issue 1–2, 2009

Editorial Board

Editor-in-Chief:
Jamie Callan
Carnegie Mellon University
callan@cmu.edu

Fabrizio Sebastiani
Consiglio Nazionale delle Ricerche
fabrizio.sebastiani@isti.cnr.it

Editors
Alan Smeaton (Dublin City University)
Andrei Z. Broder (Yahoo! Research)
Bruce Croft (University of Massachusetts, Amherst)
Charles L.A. Clarke (University of Waterloo)
Ellen Voorhees (National Institute of Standards and Technology)
Ian Ruthven (University of Strathclyde, Glasgow)
James Allan (University of Massachusetts, Amherst)
Justin Zobel (RMIT University, Melbourne)
Maarten de Rijke (University of Amsterdam)
Marcello Federico (ITC-irst)
Norbert Fuhr (University of Duisburg-Essen)
Soumen Chakrabarti (Indian Institute of Technology)
Susan Dumais (Microsoft Research)
Wei-Ying Ma (Microsoft Research Asia)
William W. Cohen (CMU)

Full text available at: http://dx.doi.org/10.1561/1500000012

Editorial Scope

Foundations and Trends® in Information Retrieval will publish survey and tutorial articles in the following topics:

- Applications of IR
- Architectures for IR
- Collaborative filtering and recommender systems
- Cross-lingual and multilingual IR
- Distributed IR and federated search
- Evaluation issues and test collections for IR
- Formal models and language models for IR
- IR on mobile platforms
- Indexing and retrieval of structured documents
- Information categorization and clustering
- Information extraction
- Information filtering and routing
- Metasearch, rank aggregation and data fusion
- Natural language processing for IR
- Performance issues for IR systems, including algorithms, data structures, optimization techniques, and scalability
- Question answering
- Summarization of single documents, multiple documents, and corpora
- Text mining
- Topic detection and tracking
- Usability, interactivity, and visualization issues in IR
- User modelling and user studies for IR
- Web search

Information for Librarians

Foundations and Trends® in Information Retrieval, 2009, Volume 3, 4 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

Full text available at: http://dx.doi.org/10.1561/1500000012

Foundations and Trends® in
Information Retrieval
Vol. 3, Nos. 1–2 (2009) 1–224
© 2009 D. Kelly
DOI: 10.1561/1500000012

now
the essence of knowledge

Methods for Evaluating Interactive Information Retrieval Systems with Users

Diane Kelly

School of Information and Library Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, dianek@email.unc.edu

Abstract

This paper provides overview and instruction regarding the evaluation of interactive information retrieval systems with users. The primary goal of this article is to catalog and compile material related to this topic into a single source. This article (1) provides historical background on the development of user-centered approaches to the evaluation of interactive information retrieval systems; (2) describes the major components of interactive information retrieval system evaluation; (3) describes different experimental designs and sampling strategies; (4) presents core instruments and data collection techniques and measures; (5) explains basic data analysis techniques; and (4) reviews and discusses previous studies. This article also discusses validity and reliability issues with respect to both measures and methods, presents background information on research ethics and discusses some ethical issues which are specific to studies of interactive information retrieval (IIR). Finally, this article concludes with a discussion of outstanding challenges and future research directions.

Full text available at: http://dx.doi.org/10.1561/1500000012

Contents

Full text available at: http://dx.doi.org/10.1561/1500000012

Full text available at: http://dx.doi.org/10.1561/1500000012

Full text available at: http://dx.doi.org/10.1561/1500000012

1

Introduction

Information retrieval (IR) has experienced huge growth in the past decade as increasing numbers and types of information systems are being developed for end-users. The incorporation of users into IR system evaluation and the study of users' information search behaviors and interactions have been identified as important concerns for IR researchers [46]. While the study of IR systems has a prescribed and dominant evaluation method that can be traced back to the Cranfield studies [54], studies of users and their interactions with information systems do not have well-established methods. For those interested in evaluating interactive information retrieval systems with users, it can be difficult to determine how to proceed from a scan of the literature since guidelines for designing and conducting such studies are for the most part missing.

In interactive information retrieval (IIR), users are typically studied along with their interactions with systems and information. While classic IR studies abstract humans out of the evaluation model, IIR focuses on users' behaviors and experiences — including physical, cognitive and affective — and the interactions that occur between users and systems, and users and information. In simple terms, classic IR evaluation asks

Full text available at: http://dx.doi.org/10.1561/1500000012

the question, does this system retrieve relevant documents? IIR evaluation asks the question, can people use this system to retrieve relevant documents? IIR studies include both system evaluations as well as more focused studies of users' information search behaviors and their interactions with systems and information. IIR is informed by many fields including traditional IR, information and library science, psychology, and human–computer interaction (HCI). IIR has often been presented more generally as a combination of IR and HCI, or as a sub-area of HCI, but Ruthven [225] argues convincingly that IIR is a distinct research area. Recently, there has been interest in HCIR, or human computer information retrieval, but this looks similar to IIR and papers about this area have not established its uniqueness (e.g., [191]).

The proposition that IR systems are fundamentally interactive and should be evaluated from the perspective of users is not new. A review of IR literature reveals that many leaders in the field were writing about and studying interactive IR systems during the early years of IR research. For instance, Salton wrote a paper entitled "*Evaluation problems in interactive information retrieval*" which was published in 1970. In this paper, Salton [229] identified user effort measures as important components of IR evaluation, including the attitudes and perceptions of users. Cleverdon et al. [55] identified presentation issues and user effort as important evaluation measures for IR systems, along with recall and precision. Tague and Schultz [259] discuss the notion of user friendliness.

Some of the first types of IR interactions were associated with relevance feedback. Looking closely at this seemingly simple type of interaction, we see the difficulties inherent in IIR studies. Assuming that users are provided with information needs, each user is likely to enter a different query, which will lead to different search results and different opportunities for relevance feedback. Each user, in turn, will provide different amounts of feedback, which will create new lists of search results. Furthermore, causes and consequences of these interactions cannot be observed easily since much of this exists in the user's head. The actions that are available for observation — querying, saving a document, providing relevance feedback — are surrogates of cognitive activities. From such observable behaviors we must *infer* cognitive

Full text available at: http://dx.doi.org/10.1561/1500000012

activity; for instance, users who save a document may do so because it changes or adds to their understanding of their information needs.

User–system interactions are influenced by a number of other factors that are neither easily observable nor measurable. Each individual user has a different cognitive composition and behavioral disposition. Users vary according to all sorts of factors including how much they know about particular topics, how motivated they are to search, how much they know about searching, how much they know about the particular work or search task they need to complete, and even their expectations and perceptions of the IIR study [139, 194]. Individual variations in these factors mean that it is difficult to create an experimental situation that all people will experience the same, which in turn, makes it difficult to establish causal relationships. Moreover, measuring these factors is not always practical since there are likely a large number of factors and no established measurement practices.

The inclusion of users into any study necessarily makes IIR, in part, a behavioral science. As a result, appropriate methods for studying interactive IR systems must unite research traditions in two sciences which can be challenging. It is also the case that different systems, interfaces and use scenarios call for different methods and metrics, and studies of behavior and interaction suggest research designs that go beyond evaluation. For these reasons, there is no strong evaluation or experimental framework for IIR evaluations as there is for IR studies. IIR researchers are able to make many choices about how to design and conduct their evaluations, but there is little guidance about how to do this.

1.1 Purpose and Scope

There is a small body of research on evaluation models, methods, and metrics for IIR, but such studies are the exception rather than the rule (e.g., [34, 149]). In contrast to other disciplines where studies of methods and experimental design comprise an important portion of the literature, there are few, if any, research programs in IIR that investigate these issues and there is little formal guidance about how to conduct such studies, despite a long-standing call for such work

Full text available at: http://dx.doi.org/10.1561/1500000012

[231]. Tague's [260, 262] work and select chapters of the edited volume by Spärck-Jones [246] provide good starting points, but these writings are 15–20-years-old. While it might be argued that Spärck-Jones' book still describes the basic methodology behind traditional IR evaluations, Tague's work, which focuses on user-centered methods, needs updating given changes in search environments, tasks, users, and measures. It is also the case that Tague's work does not discuss data analysis. One might consult a statistics textbook for this type of information, but it can sometimes be difficult to develop a solid understanding of these topics unless they are discussed within the context of one's own area of study.

The purpose of this paper is to provide a foundation on which those new to IIR can make more informed choices about how to design and conduct IIR evaluations with human subjects.[1] The primary goal is to catalog and compile material related to the IIR evaluation method into a single source. This paper proposes some guidelines for conducting one basic type of IIR study — laboratory evaluations of experimental IIR systems. This is a particular kind of IIR study, but not the only kind. This paper is also focused more on quantitative methods, rather than qualitative. This is not a statement of value or importance, but a choice necessary to maintain a reasonable scope for this paper.

This article does not prescribe a step-by-step recipe for conducting IIR evaluations. The design of IIR studies is not a linear process and it would be imprudent to present the design process in this way. Typically, method design occurs iteratively, over time. Design decisions are interdependent; each choice impacts other choices. Understanding the possibilities and limitations of different design choices help one make better decisions, but there is no single method that is appropriate for all study situations. Part of the intellectual aspects of IIR is the method design itself. Prescriptive methods imply research can only be done in

[1] The terms *user* and *subject* are often used interchangeably in published IIR studies. A distinction between these terms will be made in Section 7. Since this paper focuses primarily on laboratory evaluations, the term *subject* will be used when discussing issues related to laboratory evaluations and *user* will be used when discussing general issues related to all IIR studies. *Subject* is used to indicate a person who has been sampled from the *user* population to be included in a study.

Full text available at: http://dx.doi.org/10.1561/1500000012

one way and often prevent researchers from discovering better ways of doing things.

The focus of this paper is on text retrieval systems. The basic methodological issues presented in this paper are relevant to other types of IIR systems, but each type of IIR system will likely introduce its own special considerations and issues. Additional attention is given to the study of different types of IIR systems in the final section of this paper. Digital libraries, a specific setting where IIR occurs, are also not discussed explicitly, but again, much of the material in this paper will be relevant to those working in this area [29].

Finally, this paper surveys some of the work that has been conducted in IIR. The survey is not intended to be comprehensive. Many of the studies that are cited are used to illustrate particular evaluation issues, rather than to reflect the state-of-the-art in IIR. For a current survey of research in IIR, see Ruthven [225]. For a more historic perspective, see Belkin and Vickery [23].

1.2 Sources and Recommended Readings

A number of papers about evaluation have been consulted in the creation of this paper and have otherwise greatly influenced the content of this paper. As mentioned earlier, the works of Tague [260, 262, 263, 264] and Tague and Schultz [259] are seminal pieces. The edited volume by Spärck-Jones [246] also formed a foundation for this paper.

Other research devoted to the study and development of individual components or models for IIR evaluation have also influenced this paper. Borlund [32, 34] has contributed much to IIR evaluation with her studies of simulated information needs and evaluation measures. Haas and Kraft [115] reviewed traditional experimental designs and related these to information science research. Ingwersen and Järvelin [139] present a general discussion of methods used in information seeking and retrieval research. Finally, the TREC Interactive Track [80] and all of the participants in this Track over the years have made significant contributions to the development of an IIR evaluation framework.

Review articles have been written about many topics discussed in this paper. These articles include Sugar's [255] review of user-centered

Full text available at: http://dx.doi.org/10.1561/1500000012

perspectives in IR and Turtle et al.'s [277] review of interactive IR research as well as Ruthven's [225] more recent version. The *Annual Review of Information Science and Technology* (*ARIST*) has also published many chapters on evaluation over its 40-year history including King's [173] article on the design and evaluation of information systems,[2] Kantor's [161] review of feedback and its evaluation in IR, Rorvig's [223] review of psychometric measurement in IR, Harter and Hert's [123] review of IR system evaluation, and Wang's [290] review of methodologies and methods for user behavior research.

Several special issues of journals about evaluation of IR and IIR systems are also worth mentioning. The most current is Borlund and Ruthven's [37] special issue of *IP&M* about evaluating IIR systems. Other special issues include Dunlop et al.'s [82] special issue of *Interacting with Computers* and Harman's [120] special issue of *IP&M*, which included Robertson and Hancock-Beaulieu's [221] discussion of changes in IR evaluation as a result of new understandings of relevance, interaction and information behavior. These articles, along with Savage-Knepshield and Belkin's [240] analysis of how IR interaction has changed over time, Saracevic's [233] assessment of evaluation in IR, and Ingwersen and Järvelin's [139] book on information seeking and retrieval are great background reading for those interested in the evolution of IIR systems and evaluation.

In addition to the sources from the IIR and IR literature, a number of sources related to experimental design and statistics were instrumental in the development of this paper: Babbie [13], Cohen [56], Gravetter and Wallnau [110], Myers and Well [200], Pedhazur and Schmelkin [208], and Williams [296].

1.3 Outline of Paper

The paper begins with a description of IIR and short discussion of its history. The next section reviews general approaches to studying IIR. Although this paper focuses on laboratory evaluations, other approaches are discussed briefly. Section 5 introduces

[2] Six articles were published in *ARIST* with the title, *Design and evaluation of information systems*, during the period 1968–1975.

Full text available at: http://dx.doi.org/10.1561/1500000012

research basics — research questions, theory, hypotheses, and variables. More advanced readers might want to skip this section, although the discussion of levels of measurement is particularly important for understanding the later material on statistics. Basic experimental designs are introduced in Section 6, followed by a discussion of sampling (Section 7). Instruments and data collection techniques are then presented in Section 8, followed by a discussion of some of the more common measures used in IIR evaluation (Section 10). A lengthy section on data analysis is in Section 11; although some instruction regarding qualitative data analysis is provided, this section primarily focuses on quantitative data analysis. This presentation starts with the basics of statistical data analysis, so advanced readers might want to skim parts of this section. Discussions of validity and reliability and research ethics are in Section 12. The paper concludes with future directions and challenges in Section 14.

Full text available at: http://dx.doi.org/10.1561/1500000012

References

[1] R. P. Abelson, *Statistics as Principled Argument*. Hillsdale, NJ: Lawrence Erlbaum Publishers, 1995.

[2] E. Adar, "User 4XXXXX9: Anonymizing query logs," in *Proceedings of the Workshop on Query Log Analysis: Social and Technological Challenges, at the 16th International World Wide Web Conference*, Banff, Canada, 2007.

[3] E. Agichtein, E. Brill, and S. Dumais, "Improving web search ranking by incorporating user behavior," in *Proceedings of the 29th Annual ACM SIGIR Conference on Research and Development in Information Retrieval* (*SIGIR '06*), pp. 3–10, Seattle, WA, 2006.

[4] J. Allan, "HARD track overview in TREC 2003: High accuracy retrieval from documents," in *TREC2003, Proceedings of the 12th Text Retrieval Conference*, (E. Voorhees and L. P. Buckland, eds.), Washington, DC: GPO, 2004.

[5] J. Allan, "HARD track overview in TREC 2005: High accuracy retrieval from documents," in *TREC2005, Proceedings of the 14th Text Retrieval Conference*, (E. Voorhees and L. P. Buckland, eds.), Washington, DC: GPO, 2006.

[6] B. Allen, "Information needs: A person-in-situation approach," in *Information Seeking in Context: Proceedings of an International Conference on Research in Information Needs, Seeking and Use in Different Contexts*, (P. Vakkari, R. Savolainen, and B. Dervin, eds.), pp. 111–122, Tampere, Finland, 1997.

[7] O. Alonso, D. E. Rose, and B. Stewart, "Crowdsourcing for relevance evaluation," *SIGIR Forum*, vol. 42, pp. 10–16, 2008.

[8] R. Amdur, *Institutional Review Board Member Handbook*. Sudbury, Massachusetts: Jones and Bartlett Publishers, 2003.

[9] American Psychological Association, *Publication Manual of the American Psychological Association*. Washington, DC: APA, Fifth ed., 2001.

Full text available at: http://dx.doi.org/10.1561/1500000012

[10] E. Amitay, G. C. Murray, and J. Teevan, "Workshop on query log analysis: Social and technological challenges," in *Proceedings of the 16th International World Wide Web Conference*, Banff, Canada, 2007.

[11] P. Anick, "Using terminological feedback for web search refinement: A log based study," in *Proceedings of the 26th Annual ACM International Conference on Research and Development in Information Retrieval* (*SIGIR '03*), pp. 88–95, Toronto, CA, 2003.

[12] I. Arapakis and J. Jose, "Affective feedback: An investigation of the role of emotions during an information seeking process," in *Proceedings of the 31st Annual ACM International Conference on Research and Development in Information Retrieval* (*SIGIR '08*), pp. 395–402, Singapore, Malaysia, 2008.

[13] E. Babbie, *The Practice of Social Research*. CA, Wadsworth, 10 ed., 2004.

[14] P. Bailey, N. Craswell, I. Soboroff, P. Thomas, A. P. de Vries, and E. Yilmaz, "Relevance assessment: Are judges exchangeable and does it matter," in *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (*SIGIR '08*), pp. 667–674, Singapore, Malaysia, 2008.

[15] M. J. Bates, "Information search tactics," *Journal of the American Society for Information Science*, vol. 30, pp. 205–214, 1979.

[16] M. M. Beaulieu, "Interaction in information searching and retrieval," *Journal of Documentation*, vol. 56, pp. 431–439, 2000.

[17] M. M. Beaulieu, M. Gatford, X. Huang, S. E. Robertson, S. Walker, and P. Williams, "Okapi at TREC-5," in *Proceedings of the 5th Text Retrieval Conference* (*TREC-5*), (E. M. Voorhees and D. K. Harman, eds.), pp. 143–165, Washington, DC: GPO, 1997.

[18] B. Bederson, "Interfaces for staying in the flow," *Ubiquity*, vol. 5, 2004.

[19] N. J. Belkin, "Anomalous states of knowledge as a basis for information retrieval," *Canadian Journal of Information Science*, vol. 5, pp. 133–143, 1980.

[20] N. J. Belkin, "Helping people find what they don't know," *Communications of the ACM*, vol. 43, pp. 58–61, 2000.

[21] N. J. Belkin, A. Cabezas, C. Cool, K. Kim, K. B. Ng, S. Park, R. Pressman, S. Rieh, P. Savage, and I. Xie, "Rutgers interactive track at TREC-5," M. M. Beaulieu, M. Gatford, X. Huang, S. E. Robertson, S. Walker, and P. Williams, "Okapi at TREC-5," in *Proceedings of the 5th Text Retrieval Conference* (*TREC-5*), (E. M. Voorhees and D. K. Harman, eds.), pp. 257–265, Washington, DC: GPO, 1997, 1997.

[22] N. J. Belkin, D. Kelly, G. Kim, J.-Y. Kim, H.-J. Lee, G. Muresan, M.-C. Tang, X.-J. Yuan, and C. Cool, "Query length in interactive information retrieval," in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 205–212, Toronto, Canada, 2003.

[23] N. J. Belkin and A. Vickery, *Interaction in Information Systems: A Review of Research from Document Retrieval to Knowledge-Based Systems*. Library and Information Research Report 35: The British Library, University Press, Cambridge, 1985.

Full text available at: http://dx.doi.org/10.1561/1500000012

[24] D. J. Bell and I. Ruthven, "Searchers' assessments of task complexity for Web searching," in *Proceedings of the 26th Annual International European Conference on Information Retrieval* (*ECIR 2004*), pp. 57–71, Sunderland, UK, 2004.

[25] J. L. Bennett, "The user interface in interactive systems," *Annual Review of Information Science and Technology*, vol. 7, pp. 159–196, 1972.

[26] J. D. Bernal, "Preliminary analysis of pilot questionnaires on the use of scientific literature," *The Royal Society Scientific Information Conference*, pp. 589–637, 1948.

[27] W. H. Beyer, *Handbook of Tables for Probability and Statistics*. Cleveland, OH: Chemical Rubber Co. Publishers, 1966.

[28] M. Bilenko and R. W. White, "Mining the search trails of surfing crowds: Identifying relevant websites from user activity," in *Proceedings of the 17th International Conference on the World Wide Web* (*WWW '08*), pp. 51–60, Beijing, China, 2008.

[29] A. Blandford, A. Adams, S. Attfield, G. Buchanan, J. Gow, S. Makri, J. Rimmer, and C. Warwick, "The PRET A Rapporter framework: Evaluating digital libraries from the perspective of information work," *Information Processing and Management*, vol. 44, pp. 4–21, 2008.

[30] C. L. Borgman, "End user behavior on an online information retrieval system: A computer monitoring study," in *Proceedings of the 6th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (*SIGIR '83*), pp. 162–176, Bethesda, MD, 1983.

[31] C. L. Borgman, "All users of information retrieval systems are not created equal: An exploration into individual differences," *Information Processing and Management*, vol. 25, pp. 237–251, 1989.

[32] P. Borlund, "Experimental components for the evaluation of interactive information retrieval systems," *Journal of Documentation*, vol. 56, pp. 71–90, 2000.

[33] P. Borlund, "The concept of relevance in IR," *Journal of the American Society for Information Science*, vol. 54, pp. 913–925, 2003a.

[34] P. Borlund, "The IIR evaluation model: A framework for evaluation of interactive information retrieval systems," *Information Research*, vol. 8, p. 152, 2003b.

[35] P. Borlund and P. Ingwersen, "The development of a method for evaluating interactive information retrieval systems," *Journal of Documentation*, vol. 53, pp. 225–250, 1997.

[36] P. Borlund and P. Ingwersen, "Measure of relative relevance and ranked half-life: Performance indicators for interactive information retrieval," in *Proceedings of the 21st ACM SIGIR Conference on Research and Development of Information Retrieval* (*SIGIR '98*), pp. 324–331, Melbourne, Australia, 1998.

[37] P. Borlund and I. Ruthven, "Introduction to the special issue on evaluating interactive information retrieval systems," *Information Processing and Management*, vol. 44, pp. 1–3, 2008.

[38] P. J. Borlund, W. Schneider, M. Lalmas, A. Tombros, J. Feather, D. Kelly, A. P. de Vries, and L. Azzopardi, *Proceedings of the 2nd International Symposium on Information Interaction in Context*. London, UK, 2008.

Full text available at: http://dx.doi.org/10.1561/1500000012

[39] B. R. Boyce, C. T. Meadow, and D. H. Kraft, *Measurement in Information Science*. London, UK: Academic Press, Inc, 1994.

[40] J. Bradley, "Methodological issues and practices in qualitative research," *Library Quarterly*, vol. 63, pp. 431–449, 1993.

[41] J. Budzik and K. J. Hammond, "User interactions with everyday applications as context for just-in-time information access," in *Proceedings of the 5th International Conference on Intelligent User Interfaces* (*IUI '00*), pp. 44–51, New Orleans, LA, 2000.

[42] M. Bulmer, *Questionnaires V. 1.* Thousand Oaks, CA: Sage Publications, 2004.

[43] K. Byström, "Information and information sources in tasks of varying complexity," *Journal of the American Society for Information Science and Technology*, vol. 53, pp. 581–591, 2002.

[44] K. Byström and P. Hansen, "Conceptual framework for tasks in information studies," *Journal of the American Society for Information Science and Technology*, vol. 56, pp. 1050–1061, 2005.

[45] K. Byström and K. Järvelin, "Task complexity affects information seeking and use," *Information Processing and Management*, vol. 31, pp. 191–213, 1995.

[46] J. Callan, J. Allan, J. L. A. Clarke, S. Dumais, D. A. Evans, M. Sanderson, and C. Zhai, "Meeting of the MINDS: An information retrieval research agenda," *SIGIR Forum*, vol. 41, pp. 25–34, 2007.

[47] D. T. Campbell and J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally, 1966.

[48] R. Capra, "Studying elapsed time and task factors in re-finding electronic information," *Personal Information Management, CHI 2008 Workshop*, Florence, Italy, 2008.

[49] D. O. Case, *Looking for Information: A Survey of Research on Information Seeking, Needs and Behavior*. Lexington, KY: Academic Press, 2002.

[50] K. Charmaz, "Qualitative interviewing and grounded theory analysis," in *Handbook of Interview Research: Context and Method*, (J. F. Gubrium and J. A. Holstein, eds.), CA: Sage Publications, 2002.

[51] E. Chatman, "The impoverished life-world of outsiders," *Journal of the American Society for Information Science*, vol. 47, pp. 193–206, 1996.

[52] H. Chen, R. Wigand, and M. Nilan, "Exploring Web users' optimal flow experiences," *Information Technology and People*, vol. 13, pp. 263–281, 2000.

[53] J. P. Chin, V. A. Diehl, and K. L. Norman, "Development of an instrument measuring user satisfaction of the human–computer interface," in *Proceedings of ACM Human Factors in Computing Systems Conference* (*CHI 1988*), pp. 213–218, 1988.

[54] C. W. Cleverdon, "The Cranfield tests on index language devices," in *Readings in Information Retrieval*, (K. Spark-Jones and P. Willett, eds.), (Reprinted from *Aslib Proceedings*, pp. 173–192.) San Francisco: Morgan Kaufman Publishers, 1997/1967.

[55] C. W. Cleverdon, L. Mills, and M. Keen, *Factors Determining the Performance of Indexing Systems, vol. 1 — Design*. Cranfield, England: Aslib Cranfield Research Project, 1966.

Full text available at: http://dx.doi.org/10.1561/1500000012

[56] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates, Second ed., 1988.

[57] D. R. Compeau and C. A. Higgins, "Computer self-efficacy: Development of a measure and initial test," *MIS Quarterly*, vol. 19, pp. 189–211, 1995.

[58] C. Cool, "The concept of situation in information science," *Annual Review of Information Science and Technology*, pp. 5–42, 2001.

[59] W. S. Cooper, "Expected search length: A single measure of retrieval effectiveness based on the weak ordering action of retrieval systems," *American Documentation*, vol. 19, pp. 30–41, 1968.

[60] W. S. Cooper, "On selecting a measure of retrieval effectiveness, part 1: The "subjective" philosophy of evaluation," *Journal of the American Society for Information Science*, vol. 24, pp. 87–100, 1973.

[61] P. Cowley, J. Haack, R. Littlefield, and E. Hampson, "Glass Box: Capturing, archiving and retrieving workstation activities," in *Proceedings of the 2nd ACM Workshop on Continuous Archival and Retrieval of Personal Experiences* (*CARPE '05*), pp. 13–18, Santa Barbara, CA, 2006.

[62] F. Crestani and H. Du, "Written versus spoken queries: A qualitative and quantitative comparative analysis," *Journal of the American Society for Information Science and Technology*, vol. 57, pp. 881–890, 2006.

[63] M. Csikszentmihalyi, *Finding Flow: The Psychology of Engagement with Everyday Life*. New York: Basic Books, 1997.

[64] M. Csikszentmihalyi and R. Larson, "Validity and reliability of the experience-sampling method," *Journal of Nervous and Mental Disease*, vol. 175, pp. 526–536, 1987.

[65] M. Czerwinski, E. Horvitz, and E. Cutrell, "Subjective duration assessment: An implicit probe for software usability," in *Proceedings of IHM-HCI 2001 Conference*, pp. 167–170, Lille, France, 2001.

[66] M. Czerwinski, E. Horvitz, and S. Wilhite, "A diary study of task switching and interruptions," in *Proceedings of ACM Human Factors in Computing Systems Conference* (*CHI '04*), pp. 175–182, Vienna, Austria, 2004.

[67] N. Dahlbäck, A. Jönsson, and L. Ahrenberg, "Wizard of Oz studies: Why and how," in *Proceedings for the 1st International Conference on Intelligent User Interfaces* (*IUI '93*), pp. 193–200, Orlando, FL, 1993.

[68] H. Dang, D. Kelly, and J. Lin, "Overview of the TREC 2007 question answering track," in *TREC2007, Proceedings of the 16th Text Retrieval Conference*, (E. Voorhees and L. P. Buckland, eds.), Washington DC: GPO, 2008.

[69] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, pp. 319–340, 1989.

[70] M. De Mey, "The cognitive viewpoint: Its development and its scope," in *CC77: International Workshop on the Cognitive Viewpoint,* (M. De Mey et al., eds.), pp. xvi–xxxi, Ghent, Belgium: University of Ghent Press, 1977.

[71] S. Debowski, R. Wood, and A. Bandura, "The impact of guided exploration and enactive exploration on self-regulatory mechanisms and information acquisition through electronic enquiry," *Journal of Applied Psychology*, vol. 86, pp. 1129–1141, 2001.

[72] S. Dennis, P. Bruza, and R. McArthur, "Web searching: A process-oriented experimental study of three interactive search paradigms," *Journal of the*

Full text available at: http://dx.doi.org/10.1561/1500000012

American Society for Information Science and Technology, vol. 53, pp. 120–133, 2002.

[73] N. K. Denzin and Y. S. Lincoln, *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage Publications, 2000.

[74] B. Dervin, "From the mind's eye of the user: The sense-making qualitative-quantitative methodology," in *Qualitative Research in Information Management*, (R. Glazier, ed.), pp. 61–84, Englewood, CO: Libraries Unlimited, 1992.

[75] B. Dervin, "Given a context by any other name: Methodological tools for taming the unruly beast," in *Information Seeking in Context: Proceedings of an International Conference on Research in Information Needs, Seeking and Use in Different Contexts*, pp. 13–38, Tampere, Finland, 1996.

[76] A. Dillon, "User analysis in HCI: The historical lesson from individual differences research," *International Journal of Human–Computer Studies*, vol. 45, pp. 619–637, 1996.

[77] W. D. Dominick and W. D. Penniman, "Automated monitoring to support the analysis and evaluation of information systems," in *Proceedings of the 2nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (*SIGIR '79*), pp. 2–9, Dallas, TX, 1979.

[78] P. Dourish, "What we talk about when we talk about context," *Personal and Ubiquitous Computing*, vol. 8, pp. 19–30, 2004.

[79] S. Dumais, E. Cutrell, J. J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins, "Stuff I've Seen: A system for personal information retrieval and re-use," in *Proceedings of the ACM Conference on Research and Development in Information Retrieval* (*SIGIR '03*), pp. 72–79, Toronto, Canada, 2003.

[80] S. T. Dumais and N. J. Belkin, "The TREC interactive tracks: Putting the user into search," in *TREC: Experiment and Evaluation in Information Retrieval*, (E. M. Voorhees and D. K. Harman, eds.), pp. 123–153, Cambridge, MA: MIT Press, 2005.

[81] M. Dunlop, "Time, relevance and interaction modeling for information retrieval," in *Proceedings of the 20th ACM Conference on Research and Development in Information Retrieval* (*SIGIR '97*), pp. 206–213, Philadelphia, PA, 1997.

[82] M. D. Dunlop, C. W. Johnson, and J. Reid, "Exploring the layers of information retrieval evaluation," *Interacting with Computers*, vol. 10, pp. 225–236, 1998.

[83] D. E. Egan, J. R. Remde, L. M. Gomez, T. K. Landauer, J. Eberhardt, and C. C. Lochbaum, "Formative design-evaluation of SuperBook," *ACM Transactions on Information Systems*, vol. 7, pp. 30–57, 1989.

[84] M. Eisenberg, "Measuring relevance judgments," *Information Processing and Management*, vol. 24, pp. 373–389, 1988.

[85] D. Elgesem, "What is special about the ethical issues in online research?," *Ethics and Information Technology*, vol. 4, pp. 195–203, 2002.

[86] D. Elsweiler and I. Ruthven, "Towards task-based personal information management evaluations," in *Proceedings of the 30th ACM Conference on Research and Development in Information Retrieval* (*SIGIR '03*), pp. 22–30, Amsterdam, The Netherlands, 2007.

Full text available at: http://dx.doi.org/10.1561/1500000012

[87] K. A. Ericsson and H. A. Simon, *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: The MIT Press, Revised ed., 1993.

[88] C. H. Fenichel, "Online searching: Measures that discriminate among users with different types of experience," *Journal of the American Society for Information Science*, vol. 32, pp. 23–32, 1981.

[89] K. D. Fenstermacher and M. Ginsburg, "Client-side monitoring for Web mining," *Journal of the American Society for Information Science and Technology*, vol. 54, pp. 625–637, 2003.

[90] R. Fidel, "Online searching styles: A case-study-based model of searching behavior," *Journal of the American Society for Information Science*, vol. 35, pp. 211–221, 1984.

[91] R. Fidel, "Qualitative methods in information retrieval research," *Library and Information Science Research*, vol. 15, pp. 219–247, 1993.

[92] L. Finkelstein, E. Gabrilovich, Y. Matias, E. Rivlin, Z. Solan, G. Wolfman, and E. Ruppin, "Placing search in context: The concept revisited," *Transactions on Information Systems*, vol. 20, pp. 116–131, 2002.

[93] K. E. Fisher, S. Erdelez, and L. E. F. McKenchnie, *Theories of Information Behavior*. Medford, NJ: Information Today, 2005.

[94] D. W. Fiske, "Convergent-discriminant validation in measurements and research strategies," in *Forms of Validity in Research*, (D. Brinbirg and L. H. Kidder, eds.), pp. 77–92, San Francisco: Jossey-Bass, 1982.

[95] S. T. Fiske, "Mind the gap: In praise of informal sources of formal theory," *Personality and Social Psychology Review*, vol. 8, pp. 132–137, 2004.

[96] P. M. Fitts, R. E. Jones, and J. L. Milton, "Eye movements of aircraft pilots during instrument-landing approaches," *Aeronautical Engineering Review*, vol. 9, pp. 24–29, 1950.

[97] B. N. Flagg, *Formative Evaluation for Educational Technologies*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1990.

[98] S. Flicker, D. Haans, and H. Skinner, "Ethical dilemmas in research on Internet communities," *Qualitative Health Research*, vol. 14, pp. 124–134, 2004.

[99] N. Ford, D. Miller, and N. Moss, "The role of individual differences in Internet searching: An empirical study," *Journal of the American Society for Information Science and Technology*, vol. 52, pp. 1049–1066, 2001.

[100] N. Ford, D. Miller, and N. Moss, "Web search strategies and human individual differences: Cognitive and demographic factors, Internet attitudes and approaches," *Journal of the American Society for Information Science and Technology*, vol. 56, pp. 741–756, 2005.

[101] N. Ford, T. D. Wilson, A. Foster, D. Ellis, and A. Spink, "Information seeking and mediated searching. Part 4: Cognitive styles in information seeking," *Journal of the American Society for Information Science and Technology*, vol. 53, pp. 728–735, 2002.

[102] J. Foster, "Collaborative information seeking and retrieval," *Annual Review of Information Science and Technology*, vol. 40, pp. 329–356, 2006.

[103] L. Freund and E. Toms, "Revisiting informativeness as a process measure for information interaction," in *Proceedings of the Web Information-Seeking and Interaction (WISI) Workshop at the 30th Annual International ACM SIGIR*

Full text available at: http://dx.doi.org/10.1561/1500000012

Conference on Research and Development in Information Retrieval (*SIGIR '07*), pp. 33–36, Amsterdam, The Netherlands, 2007.

[104] G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais, "The vocabulary problem in human–system communication: An analysis and a solution," *Communications of the ACM*, vol. 30, pp. 964–971, 1987.

[105] R. M. Furr and V. R. Bacharach, *Psychometrics: An Introduction*. Sage Publications, Inc, 2007.

[106] J. A. Ghani and S. P. Deshpande, "Task characteristics and the experience of optimal flow in Human–Computer interaction," *Journal of Psychology*, vol. 128, pp. 381–391, 1994.

[107] J. A. Ghani, R. Supnick, and P. Rooney, "The experience of flow in computer-mediated and in face-to-face groups," in *Proceedings of International Conference on Information Systems* (*ICIS 1991*), pp. 229–237, New York, NY, 1991.

[108] B. G. Glaser and A. L. Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine, 1967.

[109] A. Göker and H. Myrhaug, "Evaluation of a mobile information system in context," *Information Processing and Management*, vol. 44, pp. 39–65, 2008.

[110] F. J. Gravetter and L. B. Wallnau, *Statistics for the Behavioral Sciences*. Thomson Learning, Fifth ed., 1999.

[111] C. Grimes, D. Tang, and D. M. Russell, "Query logs alone are not enough," in *Proceedings of the Workshop on Query Log Analysis: Social and Technology Challenges at the 16th International World Wide Web Conference*, Banff, Canada, 2007.

[112] Z. Guan, S. Lee, E. Cuddihy, and J. Ramey, "The validity of the stimulated retrospective think-aloud method as measured by eye tracking," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 1253–1262, Montreal, Canada, 2006.

[113] J. P. Guilford, *Fundamental Statistics in Psychology and Education*. New York: McGraw Hill, 1956.

[114] J. Gwizdka, "Revisiting search task difficulty: Behavioral and individual difference measures," *Proceedings of the 71th Annual Meeting of the American Society for Information Science and Technology* (*ASIS and T*), 2008.

[115] D. F. Haas and D. H. Kraft, "Experimental and quasi-experimental designs for research in information science," *Information Processing and Management*, vol. 20, pp. 229–237, 1984.

[116] K. Halttunen and K. Järvelin, "Assessing learning outcomes in two information retrieval learning environments," *Information Processing and Management*, vol. 41, pp. 949–972, 2005.

[117] P. A. Hancock and N. Meshkati, *Human Mental Workload*. The Netherlands: Elsevier Science Publishers, 1988.

[118] S. Hansell, AOL removes search data on vase group of Web users, New York Times, Friday, March 14, 2008. Business Section. http://query.nytimes.com/gst/fullpage.html?res=9504E5D81E3FF93BA3575BC0A9609C8B63, 2006.

[119] S. Harada, M. Naaman, Y. J. Song, Q.-Y. Wang, and A. Paepcke, "Lost in memories: Interacting with photo collections on PDAs," in *Proceedings of*

Full text available at: http://dx.doi.org/10.1561/1500000012

the 4th ACM/IEEE-CS Joint Conference on Digital Libraries (*JCDL '04*), pp. 325–333, Tuscon, AZ, 2004.

[120] D. K. Harman, "Introduction to special issue on evaluation issues in information retrieval," *Information Processing and Management*, vol. 28, pp. 439–440, 1992.

[121] D. K. Harman, "The TREC test collection," in *TREC: Experiment and Evaluation in Information Retrieval*, (E. M. Voorhees and D. K. Harman, eds.), pp. 21–52, Cambridge, MA: MIT Press, 2005.

[122] S. G. Hart and L. E. Staveland, "Development of a NASA-TLX (task load index): Results of empirical and theoretical research," in *Human Mental Workload*, (P. Hancock and N. Meshkati, eds.), pp. 139–183, The Netherlands: Elsevier Science Publishers, 1988.

[123] S. P. Harter and C. A. Hert, "Evaluation of information retrieval systems: Approaches, issues and methods," *Annual Review of Information Science and Technology*, vol. 32, pp. 3–94, 1997.

[124] M. Hassenzahl, A. Platz, M. Burmester, and K. Lehner, "Hedonic and ergonomic quality aspects determine a software's appeal," *Proceedings of the ACM Conference on Human Factors in Computing Systems* (*CHI '00*), pp. 201–208, 2000.

[125] D. Hawking, P. Bailey, and N. Craswell, "Efficient and flexible search using text and metadata," *CSIRO Mathematical and Information Sciences Tech Report, 2000-83*, available online at http://es.csiro.au/pubs/hawking_tr00b.pdf, 2000.

[126] M. D. Heine, "Simulation, and simulation experiments," in *Information Retrieval Experiment*, (K. Spärck-Jones, ed.), pp. 179–198, London, UK: Butterworths and Co. Ltd, 1981.

[127] W. Hersh, "Evaluating interactive question answering," in *Advances in Open Domain Question Answering*, (T. Strzalkowski and S. Harabagiu, eds.), pp. 431–455, Dordrecht, The Netherlands: Springer, 2006.

[128] W. Hersh and P. Over, "Introduction to a special issue on interactivity at the Text Retrieval Conference (TREC)," *Information Processing and Management*, vol. 37, pp. 365–367, 2001.

[129] W. Hersh, A. Turpin, S. Price, B. Chan, D. Kramer, L. Sacherek, and D. Olson, "Do batch and user evaluations give the same results?," in *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (*SIGIR '00*), pp. 17–24, Athens, Greece, 2000.

[130] W. R. Hersh, D. L. Elliot, D. H. Hickam, S. L. Wolf, A. Molnar, and C. Leichtenstein, "Towards new measures of information retrieval evaluation," in *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (*SIGIR '95*), pp. 164–170, Seattle, WA, 1995.

[131] S. Hirsh and J. Dinkelacker, "Seeking information in order to produce information: An empirical study at Hewlett Packard Labs," *Journal of the American Society for Information Science and Technology*, vol. 55, pp. 807–817, 2004.

Full text available at: http://dx.doi.org/10.1561/1500000012

[132] K. Hornbaek, "Current practice in measuring usability: Challenges to usability studies and research," *International Journal of Human–Computer Studies*, vol. 64, pp. 79–102, 2005.

[133] K. Hornbæk and E. L.-C. Law, "Meta-analysis of correlations among usability measures," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 617–626, San Jose, CA, 2007.

[134] I. Hsieh-Yee, "Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers," *Journal of the American Society for Information Science*, vol. 44, pp. 161–174, 1993.

[135] M. Huang and H. Wang, "The influence of document presentation order and number of documents judged on users' judgments of relevance," *Journal of the American Society for Information Science*, vol. 55, pp. 970–979, 2004.

[136] G. Iachello and J. Hong, "End-user privacy in human–computer interaction," *Foundations and Trends in Human–Computer Interaction*, vol. 1, pp. 1–137, 2007.

[137] P. Ingwersen, *Information Retrieval Interaction*. London: Taylor Graham, 1992.

[138] P. Ingwersen, "Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory," *Journal of Documentation*, vol. 52, pp. 3–50, 1996.

[139] P. Ingwersen and K. Järvelin, *The Turn: Integration of Information Seeking and Retrieval in Context*. Dordrecht, The Netherlands: Springer, 2005.

[140] P. Ingwersen and P. Willett, "An introduction to algorithmic and cognitive approaches for information retrieval," *Libri*, vol. 45, pp. 160–177, 1995.

[141] International Standards Office, *Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs): Part II, Guidance on Usability* (ISO 9241-11:1998). 1998.

[142] E. Jacob, "Qualitative research traditions: A review," *Review of Educational Research*, vol. 57, pp. 1–50, 1987.

[143] R. J. K. Jacob and K. S. Karn, "Eye tracking in human–Computer interaction and usability research: Ready to deliver the promises (section commentary)," in *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, (J. Hyona, R. Radach, and H. Deubel, eds.), pp. 573–605, Amsterdam, Elsevier Science, 2003.

[144] B. J. Jansen, "Search log analysis: What it is, what's been done, how to do it," *Library and Information Science Research*, vol. 28, pp. 407–432, 2006.

[145] B. J. Jansen, R. Ramadoss, M. Zhang, and N. Zang, "Wrapper: An application for evaluating exploratory searching outside of the lab," in *Workshop on Evaluating Exploratory Search Systems at the 29th Annual International ACM SIGIR Conference on Research and Development on Information Retrieval (SIGIR '06)*, Seattle, WA, 2006.

[146] B. J. Jansen and A. Spink, "How are we searching the World Wide Web? A comparison of nine search engine transaction logs," *Information Processing and Management*, vol. 42, pp. 248–263, 2005.

Full text available at: http://dx.doi.org/10.1561/1500000012

[147] K. Järvelin, "An analysis of two approaches in information retrieval: From frameworks to study designs," *Journal of the American Society for Information Science and Technology*, vol. 58, pp. 971–986, 2007.

[148] K. Järvelin and J. Kekäläinen, "IR evaluation methods for retrieving highly relevant documents," in *Proceedings of the 23rd ACM SIGIR Conference on Research and Development of Information Retrieval* (*SIGIR '00*), pp. 41–48, Athens, Greece, 2000.

[149] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques," *ACM Transactions on Information Systems* (*TOIS*), vol. 20, pp. 422–446, 2002.

[150] K. Järvelin, S. L. Price, L. M. L. Delcambre, and M. L. Nielsen, "Discounted cumulated gain based evaluation of multiple-query IR sessions," in *Proceedings of the 30th European Conference on Information Retrieval*, Glasgow, Scotland, 2008.

[151] H. R. Jex, "Measuring mental workload: Problems, process and promises," in *Human Mental Workload*, (P. Hancock and N. Meshkati, eds.), pp. 5–39, The Netherlands: Elsevier Science Publishers, 1988.

[152] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay, "Accurately interpreting clickthrough data as implicit feedback," in *Proceedings of the 28th Annual ACM International Conference on Research and Development in Information Retrieval* (*SIGIR '05*), pp. 154–161, Salvador, Brazil, 2005.

[153] H. Joho, D. Hannah, and J. M. Jose, "Comparing collaborative and independent search in a recall-oriented task," in *Proceedings of the 2nd IIiX Symposium on Information Interaction in Context*, pp. 89–96, London, UK, 2008.

[154] H. Joho and J. M. Jose, "Effectiveness of additional representations for the search result presentation on the Web," *Information Processing and Management*, vol. 44, pp. 226–241, 2008.

[155] W. Jones and J. Teevan, *Personal Information Management*. Seattle, WA: University of Washington Press, 2007.

[156] C.-A. Julien, J. E. Leide, and F. Bouthillier, "Controlled user evaluations of information visualization interfaces for text retrieval: Literature review and meta-analysis," *Journal of the American Society for Information Science and Technology*, vol. 59, pp. 1012–1024, 2008.

[157] H. Julien, L. E. F. McKechnie, and S. Hart, "Affective issues in library and information science systems work: A content analysis," *Library and Information Science Research*, vol. 27, pp. 453–466, 2005.

[158] M. Käki and A. Aula, "Controlling the complexity in comparing search user interfaces via user studies," *Information Processing and Management*, vol. 44, pp. 82–91, 2008.

[159] J. Kalgren and K. Franzen, "Verbosity and interface design," Retrieved on 01 February 2008 at http://www.ling.su.se/staff/franzen/irinterface.html, 1997.

[160] P. Kantor, G. Kazai, N. Milic-Frayling, and R. Wilkinson, "Proceedings of the 2008 ACM workshop on research advances in large digital book repositories," in *Proceedings of the Conference on Information and Knowledge Management* (*CIKM '08*), Napa, CA, 2008.

Full text available at: http://dx.doi.org/10.1561/1500000012

[161] P. B. Kantor, "Evaluation of and feedback in information storage and retrieval systems," *Annual Review of Information Science and Technology*, vol. 17, pp. 99–120, 1982.

[162] J. Kekäläinen and K. Järvelin, "Using graded relevance assessments in IR evaluation," *Journal of the American Society for Information Science and Technology*, vol. 53, pp. 1120–1129, 2002.

[163] M. Kellar, C. Watters, and M. Shepherd, "A field study characterizing Web-based information-seeking tasks," *Journal of the American Society for Information Science and Technology*, vol. 58, pp. 999–1018, 2007.

[164] D. Kelly and X. Fu, "Elicitation of term relevance feedback: An investigation of term source and context," in *Proceedings of the 29th Annual ACM International Conference on Research and Development in Information Retrieval* (*SIGIR '06*), pp. 453–460, Seattle, WA, 2006.

[165] D. Kelly, J. Harper, and B. Landau, "Questionnaire mode effects in interactive information retrieval experiments," *Information Processing and Management*, 2008.

[166] D. Kelly, P. B. Kantor, E. L. Morse, J. Scholtz, and Y. Sun, "User-centered evaluation of interactive question answering systems," in *Proceedings of the Workshop on Interactive Question Answering at the Human Language Technology Conference* (*HLT-NAACL '06*), pp. New York, NY, 2006.

[167] D. Kelly and J. Lin, "Overview of the TREC 2006 ciQA task," *SIGIR Forum*, vol. 41, pp. 107–116, 2007.

[168] D. Kelly, C. Shah, C. R. Sugimoto, E. W. Bailey, R. A. Clemens, A. K. Irvine, N. A. Johnson, W. Ke, S. Oh, A. Poljakova, M. A. Rodriguez, M. G. van Noord, and Y. Zhang, "Effects of performance feedback on users' evaluations of an interactive IR system," in *Proceedings of the 2nd Symposium on Information Interaction in Context* (*IIiX*), pp. 75–82, London, UK, 2008.

[169] D. Kelly, N. Wacholder, R. Rittman, Y. Sun, P. Kantor, S. Small, and T. Strzalkowski, "Using interview data to identify evaluation criteria for interactive, analytical question answering systems," *Journal of the American Society for Information Science and Technology*, vol. 58, pp. 1032–1043, 2007.

[170] J. Kim, "Task as a predictable indicator for information seeking behavior on the Web," Ph.D. Dissertation, Rutgers University, 2006.

[171] K.-S. Kim and B. Allen, "Cognitive and task influences on Web searching behavior," *Journal of the American Society for Information Science and Technology*, vol. 53, pp. 109–119, 2001.

[172] S. Kim and D. Soergel, "Selecting and measuring task characteristics as independent variables," *Proceedings of the American Society for Information Science and Technology Conference*, vol. 42, 2006.

[173] D. W. King, "Design and evaluation of information systems," *Annual Review of Information Science and Technology*, vol. 3, pp. 61–103, 1968.

[174] C. C. Kuhlthau, *Seeking Meaning: A Process Approach to Library and Information Services*. Norwood, NJ: Ablex, 1993.

[175] E. Lagergren and P. Over, "Comparing interactive information retrieval systems across sites: The TREC-6 interactive track matrix experiment," in *Proceedings of the 21st Annual International ACM SIGIR Conference on*

Full text available at: http://dx.doi.org/10.1561/1500000012

Research and Development in Information Retrieval, pp. 164–172, Melbourne, Australia, 1998.

[176] B. Larsen, S. Malik, and A. Tombros, "The interactive track at INEX2005," in *INEX 2005*, (N. Fuhr, M. Lalmas, S. Malik, and G. Kazai, eds.), pp. 398–410, Berlin: Springer, 2006.

[177] Y. Li and N. J. Belkin, "A faceted approach to conceptualizing tasks information seeking," *Information Processing and Management*, vol. 44, pp. 1822–1837, 2008.

[178] R. Likert, "A technique for the measurement of attitudes," *Archives of Psychology*, vol. 140, pp. 1–55, 1932.

[179] J. Lin, D. Quan, V. Sinha, K. Bakshi, D. Huynh, B. Katz, and D. R. Karger, "What makes a good answer? The role of context in question answering," in *Proceedings of the 9th IFIP TC13 International Conference on Human–Computer Interaction* (*INTERACT 2003*), (M. Rauterberg, M. Menozzi, and J. Wesson, eds.), Zurich, Switzerland, 2003.

[180] J. Lin and M. Smucker, "How do users find things with PubMed? Towards automatic utility evaluation with user simulations," in *Proceedings of the 31st International ACM SIGIR Conference on Research and Development in Information Retrieval* (*SIGIR 08*), pp. 19–26, Singapore, Malaysia, 2008.

[181] S. J. Lin and N. J. Belkin, "Modeling multiple information seeking episodes," in *Proceedings of the Annual Conference of the American Society for Information Science* (*ASIS '00*), pp. 133–146, USA, 2000.

[182] S. Littlejohn, *Theories of Human Communication*. Belmont, CA: Wadsworth, 1992.

[183] I. Lopatovska and H. B. Mokros, "Willingness to pay and experienced utility as measures of affective value of information objects: Users' accounts," *Information Processing and Management*, vol. 44, pp. 92–104, 2008.

[184] L. Lorigo, M. Haridasan, H. Brynjarsdottir, L. Xia, T. Joachims, G. Gay, L. Granka, F. Pellacini, and B. Pan, "Eye tracking and online search: Lessons learned and challenges ahead," *Journal of the American Society for Information Science and Technology*, vol. 59, pp. 1041–1052, 2008.

[185] L. Lorigo, B. Pan, H. Hembrooke, T. Joachims, L. Granka, and G. Gay, "The influence of task and gender on search and evaluation behavior using Google," *Information Processing and Management*, vol. 42, pp. 1123–1131, 2006.

[186] R. M. Losee, "Evaluating retrieval performance given database and query characteristics: Analytical determination of performance surfaces," *Journal of the American Society for Information Science*, vol. 47, pp. 95–105, 1996.

[187] H. P. Luhn, "A business intelligence system," in *H. P. Luhn: Pioneer of Information Science. Selected Works*, (C. K. Shultz, ed.), pp. 132–139, NY: Spartan Books, 1958.

[188] A. M. Lund, "Measuring usability with the USE Questionnaire," *Usability and User Experience*, vol. 8, Available online: http://www.stcsig.org/ usability/newsletter/0110_measuring_with_use.html, 2001.

Full text available at: http://dx.doi.org/10.1561/1500000012

[189] W. E. Mackay, "Ethics, lies and videotape," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 138–145, Denver, CO, 1995.

[190] G. Marchionini, "Exploratory search: From finding to understanding," *Communications of the ACM*, vol. 49, pp. 41–46, 2006a.

[191] G. Marchionini, "Toward human–Computer information retrieval," *Bulletin of the American Society for Information Science*, online: http://www.asis.org/Bulletin/Jun-06/marchionini.html (retrieved November 7, 2008), 2006b.

[192] G. Marchionini and G. Crane, "Evaluating hypermedia and learning: Methods and results from the Perseus Project," *ACM Transactions on Information Systems*, vol. 12, pp. 5–34, 1994.

[193] J. A. Maxwell, *Qualitative Research Design: An Interactive Approach*. CA: Sage Publications, 1996.

[194] D. A. Michel, "What is used during cognitive processing in information retrieval and library searching? Eleven sources of search information," *Journal of the American Society for Information Science*, vol. 45, pp. 498–514, 1994.

[195] M. B. Miles and A. M. Huberman, *Qualitative Data Analysis: A Sourcebook of New Methods*. Newbury Park: Sage, 1984.

[196] L. I. Millett, B. Friedman, and E. Felten, "Cookies and web browser design: Toward realizing informed consent online," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 46–52, Seattle, WA, 2001.

[197] S. Mizzaro, "Relevance: The whole history," *Journal of the American Society for Information Science*, vol. 48, pp. 810–832, 1997.

[198] J. Morahan-Martin, "Males, females, and the Internet," in *Psychology and the Internet: Intrapersonal, Interpersonal, and Transpersonal Implications*, (J. Gackenback, ed.), pp. 169–197, San Diego: Academic Press, 1998.

[199] M. R. Morris and E. Horvitz, "SearchTogether: An interface for collaborative Web search," in *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology (UIST '07)*, pp. 3–12, New York, NY, 2007.

[200] J. L. Myers and A. D. Well, *Research Design and Statistical Analysis*. Mahway, NJ: Lawrence Erlbaum Associates Inc., Publishers, Second ed., 2003.

[201] K. A. Neuendorf, *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage Publications, 2002.

[202] J. Nielsen, "Usability 101: Introduction to usability," *Jakob Nielsen's Alertbox*, Retrieved May 05, 2008 from http://www.useit.com/alertbox/20030825.html, 2003.

[203] J. Nielsen and J. Levy, "Measuring usability: Preference vs performance," *Communications of the ACM*, vol. 37, pp. 66–75, 1994.

[204] D. Oard, "Evaluating interactive cross-language information retrieval: Document selection," in *Proceedings of the 1st Cross-Language Evaluation Forum*, Lisbon, Portugal, 2000.

[205] H. O'Brien and E. Toms, "What is user engagement? A conceptual framework for defining user engagement with technology," *Journal of the American Society for Information Science and Technology*, vol. 59, pp. 938–955, 2008.

Full text available at: http://dx.doi.org/10.1561/1500000012

[206] H. L. O'Brien, E. G. Toms, E. K. Kelloway, and E. Kelley, "Developing and evaluating a reliable measure of user engagement," in *Proceedings of the American Society for Information Science and Technology*, Columbus, Ohio, 2008.

[207] S. L. Payne, *The Art of Asking Questions*. Princeton, NJ: Princeton University Press, 1951.

[208] E. J. Pedhazur and L. P. Schmelkin, *Measurement, Design and Analysis: An Integrated Approach*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1991.

[209] T. A. Peters, "The history and development of transaction log analysis," *Library Hi Tech*, vol. 11, pp. 41–66, 1993.

[210] T. A. Peters, M. Kurth, P. Flaherty, B. Sandore, and N. K. Kaske, "An introduction to the special section on transaction log analysis," *Library Hi Tech*, vol. 11, pp. 38–40, 1993.

[211] D. Petrelli, "On the role of user-centred evaluation in the advancement of interactive information retrieval," *Information Processing and Management*, vol. 44, pp. 22–38, 2008.

[212] R. W. Picard, *Affective Computing*. Cambridge, MA: MIT Press, 1997.

[213] E. M. Pilke, "Flow experiences in information technology use," *International Journal of Human–Computer Studies*, vol. 61, pp. 347–357, 2004.

[214] D. J. Pittenger, "Internet research: An opportunity to revisit classic ethical problems in behavioral research," *Ethics and Behavior*, vol. 13, pp. 45–60, 2003.

[215] P. M. Podsakoff, S. B. MacKenzie, J.-Y. Lee, and N. P. Podsakoff, "Common method biases in behavioral research: A critical review of the literature and recommended remedies," *Journal of Applied Psychology*, vol. 88, pp. 879–903, 2003.

[216] R. J. Riding and I. Cheema, "Cognitive styles — An overview and integration," *Education Psychology*, vol. 11, pp. 193–215, 1991.

[217] S. Y. Rieh, "Judgment of information quality and cognitive authority in the Web," *Journal of the American Society for Information Science and Technology*, vol. 53, pp. 145–161, 2002.

[218] S. E. Robertson, "The probability ranking principle in IR," *Journal of Documentation*, vol. 33, pp. 294–304, 1977.

[219] S. E. Robertson, "On GMAP — And other transformations," in *Proceedings of the 15th ACM international Conference on information and Knowledge Management* (*CIKM'06*), pp. 78–83, Arlington, VA, 2006.

[220] S. E. Robertson, "On the history of evaluation in IR," *Journal of Information Science*, vol. 34, pp. 439–456, 2008.

[221] S. E. Robertson and M. M. Hancock-Beaulieu, "On the evaluation of IR systems," *Information Processing and Management*, vol. 28, pp. 457–466, 1992.

[222] K. Rodden and X. Fu, "Exploring how mouse movements relate to eye movements on Web search results pages," in *Proceedings of the Web Information-Seeking and Interaction* (*WISI*) *Workshop at the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (*SIGIR '07*), pp. 29–32, Amsterdam, The Netherlands, 2007.

[223] M. E. Rorvig, "Psychometric measurement and information retrieval," *Annual Review of Information Science and Technology*, vol. 23, pp. 157–189, 1988.

Full text available at: http://dx.doi.org/10.1561/1500000012

[224] I. Ruthven, "Integrating approaches to relevance," in *New Directions in Cognitive Information Retrieval*, (A. Spink and C. Cole, eds.), pp. 61–80, Netherlands: Springer, 2005.

[225] I. Ruthven, "Interactive information retrieval," *Annual Review of Information Science and Technology*, vol. 42, pp. 43–91, 2008.

[226] I. Ruthven, M. Baillie, and D. Elsweiler, "The relative effects of knowledge, interest and confidence in assessing relevance," *Journal of Documentation*, vol. 63, pp. 482–504, 2007.

[227] I. Ruthven, P. Borlund, P. Ingwersen, N. J. Belkin, A. Tombros, and P. Vakkari in *Proceedings of the 1st International Conference on Information Interaction in Context*, Copenhagen, Denmark, 2006.

[228] K. J. Ryan, J. V. Brady, R. E. Cooke, D. I. Height, A. R. Jonsen, P. King, K. Lebacqz, D. W. Louisell, D. Seldin, E. Stellar, and R. H. Turtle, "The Belmont Report," Available at http://www.hhs.gov/ohrp/humansubjects/ guidance/belmont.htm, 1979.

[229] G. Salton, "Evaluation problems in interactive information retrieval," *Information Storage and Retrieval*, vol. 6, pp. 29–44, 1970.

[230] G. Salton, "The state of retrieval system evaluation," *Information Processing and Management*, vol. 28, pp. 441–449, 1992.

[231] T. Saracevic, "Quo vadis test and evaluation," in *Proceedings of the Annual Meeting of the American Documentation Institute, 4*, pp. 100–104, New York, NY, 1967.

[232] T. Saracevic, "Relevance: A review of and a framework for the thinking on the notion in information science," *Journal of the American Society for Information Science*, vol. 26, pp. 321–343, 1975.

[233] T. Saracevic, "Evaluation of evaluation in information retrieval," in *Proceedings of the 18th Annual ACM SIGIR Conference on Research and Development of Information Retrieval*, pp. 138–146, Seattle, WA, 1995.

[234] T. Saracevic, "The stratified model of information retrieval interaction: Extension and applications," *Proceedings of the American Society for Information Science Conference*, vol. 34, pp. 313–327, 1997.

[235] T. Saracevic, "Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance," *Journal of the American Society for Information Science and Technology*, vol. 58, pp. 1915–1933, 2007a.

[236] T. Saracevic, "Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance," *Journal of the American Society for Information Science and Technology*, vol. 58, pp. 2126–2144, 2007b.

[237] T. Saracevic and P. B. Kantor, "A study of information seeking and retrieving. II. Users, questions and effectiveness," *Journal of the American Society for Information Science*, vol. 39, pp. 177–196, 1988a.

[238] T. Saracevic and P. B. Kantor, "A study of information seeking and retrieving. III. Searchers, searches and overlap," *Journal of the American Society for Information Science*, vol. 39, pp. 197–216, 1988b.

[239] T. Saracevic, P. B. Kantor, A. Y. Chamis, and D. Trivison, "A study of information seeking and retrieving: Part 1, background and methodology,"

Full text available at: http://dx.doi.org/10.1561/1500000012

Journal of the American Society for Information Science, vol. 39, pp. 161–176, 1988.

[240] P. A. Savage-Knepshield and N. J. Belkin, "Interaction in information retrieval: Trends over time," *Journal of the American Society for Information Science*, vol. 50, pp. 1067–1082, 1999.

[241] R. Savolainen, "Everyday life information seeking: Approaching information seeking in the context of way of life," *Library and Information Science Research*, vol. 17, pp. 259–294, 1995.

[242] R. Siatri, "The evolution of user studies," *Libri*, vol. 49, pp. 132–141, 1999.

[243] C. L. Smith and P. B. Kantor, "User adaptation: Good results from poor systems," in *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (*SIGIR '08*), pp. 147–154, Singapore, 2008.

[244] P. Solomon, "Discovering information in context," *Annual Review of Information Science and Technology*, vol. 36, pp. 229–264, 2002.

[245] E. Sormunen, "Liberal relevance criteria of TREC: Counting on negligible documents?," in *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (*SIGIR '02*), pp. 324–330, Tampere, Finland, 2002.

[246] K. Spärck-Jones, *Information Retrieval Experiment*. London, UK: Butterworths and Co. Ltd, 1981.

[247] A. Spink, "Multiple search session model of end-user behavior: An exploratory study," *Journal of the American Society for Information Science*, vol. 47, pp. 603–609, 1996.

[248] A. Spink, "Study of interactive feedback during mediated information retrieval," *Journal of the American Society for Information Science*, vol. 48, pp. 382–394, 1997.

[249] A. Spink and H. Greisdorf, "Regions and levels: Measuring and mapping users' relevance judgments," *Journal of the American Society for Information Science and Technology*, vol. 52, pp. 161–173, 2001.

[250] A. Spink and R. M. Losee, "Feedback in information retrieval," *Annual Review of Information Science and Technology*, vol. 31, pp. 33–78, 1996.

[251] A. Spink, H. C. Ozmutlu, and S. Ozmutlu, "Multitasking information seeking and searching processes," *Journal of the American Society for Information Science and Technology*, vol. 53, pp. 639–652, 2002.

[252] S. R. Stern, "Encountering distressing information in online research: A consideration of legal and ethical responsibilities," *New Media and Society*, vol. 5, pp. 249–266, 2003.

[253] A. Strauss and J. Corbin, *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. CA: Sage Publications, 1990.

[254] L. T. Su, "Evaluation measures for interactive information retrieval," *Information Processing and Management*, vol. 28, pp. 503–516, 1992.

[255] W. Sugar, "User-centered perspectives of information retrieval research and analysis methods," *Annual Review of Information Science and Technology*, vol. 30, pp. 77–109, 1995.

[256] SUMI Questionnaire. (retrieved November 05, 2008), http://www.ucc.ie/hfrg/questionnaires/sumi/index.html.

Full text available at: http://dx.doi.org/10.1561/1500000012

[257] Y. Sun and P. Kantor, "Cross-evaluation: A new model for information system evaluation," *Journal of American Society for Information Science and Technology*, vol. 56, pp. 614–628, 2006.

[258] J. Tague, "Informativeness as an ordinal utility function for information retrieval," *SIGIR Forum*, vol. 21, pp. 10–17, 1987.

[259] J. Tague and R. Schultz, "Evaluation of the user interface in an information retrieval system: A model," *Information Processing and Management*, vol. 25, pp. 377–389, 1988.

[260] J. M. Tague, "The pragmatics of information retrieval experimentation," in *Information Retrieval Experiment*, (K. S. Jones, ed.), pp. 59–104, London, UK: Butterworths and Co. Ltd, 1981.

[261] J. M. Tague-Sutcliffe, "The pragmatics of information retrieval experimentation, revisted," *Information Processing and Management*, vol. 28, pp. 467–490, 1992a.

[262] J. M. Tague-Sutcliffe, "Measuring the informativeness of a retrieval process," in *Proceedings of the 15th Annual ACM International Conference on Research and Development in Information Retrieval* (*SIGIR '92*), pp. 23–36, Copenhagen, Denmark, 1992b.

[263] J. M. Tague-Sutcliffe, *Measuring Information: An Information Services Perspective*. San Diego, California: Academic Press, 1995.

[264] J. M. Tague-Sutcliffe, "Some perspectives on the evaluation of information retrieval systems," *Journal of the American Society for Information Science*, vol. 47, pp. 1–3, 1996.

[265] R. Tang, M. Shaw, and J. L. Vevea, "Towards the identification of the optimal number of relevance categories," *Journal of the American Society for Information Science*, vol. 50, pp. 254–264, 1999.

[266] A. R. Taylor, C. Cool, N. J. Belkin, and W. J. Amadio, "Relationships between categories of relevance criteria and stage in task completion," *Information Processing and Management*, vol. 43, pp. 1071–1084, 2007.

[267] R. S. Taylor, "Question negotiation and information seeking in libraries," *College and Research Libraries*, vol. 29, pp. 178–194, 1968.

[268] J. Teevan, C. Alvarado, M. S. Ackerman, and D. R. Karger, "The perfect search engine is not enough: A study of orienteering behavior in directed search," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (*SIGCHI '04*), pp. 415–422, Vienna, Austria, 2004.

[269] J. Teevan, S. T. Dumais, and E. Horvitz, "Personalizing search via automated analysis of interests and activities," in *Proceedings of the 28th Annual ACM International Conference on Research and Development in Information Retrieval* (*SIGIR '05*), pp. 449–456, Salvador, Brazil, 2005.

[270] P. Thomas and D. Hawking, "Evaluation by comparing result sets in context," in *Proceedings of the 15th Annual Conference on Information and Knowledge Management* (*CIKM '06*), pp. 94–101, Arlington, VA, 2006.

[271] A. Tombros, I. Ruthven, and J. M. Jose, "How users assess Web pages for information seeking," *Journal of the American Society for Information Science and Technology*, vol. 56, pp. 327–344, 2005.

Full text available at: http://dx.doi.org/10.1561/1500000012

[272] E. G. Toms, L. Freund, and C. Li, "W*i*IRE: The Web interactive information retrieval experimentation system prototype," *Information Processing and Management*, vol. 40, pp. 655–675, 2004.

[273] E. G. Toms, H. L. O'Brien, T. Mackenzie, C. Jordan, L. Freund, S. Toze, E. Dawe, and A. MacNutt, "Task effects on interactive search: The query factor," *Proceedings of INEX 2007*, pp. 359–372, 2007.

[274] R. Tourangeau, L. J. Rips, and K. Rasinski, *The Psychology of Survey Response*. New York, NY: Cambridge University Press, 2000.

[275] A. Turpin and W. Hersh, "Why batch and user evaluations do not give the same results," in *Proceedings of the 24th Annual ACM International Conference on Research and Development in Information Retrieval* (*SIGIR '01*), pp. 225–231, New Orleans, LA, 2001.

[276] A. Turpin and F. Scholer, "User performance versus precision measures for simple search tasks," in *Proceedings of the 29th Annual ACM International Conference on Research and Development in Information Retrieval* (*SIGIR '06*), pp. 11–18, Seattle, WA, 2006.

[277] H. Turtle, W. D. Penniman, and T. Hickey, "Data entry/display devices for interactive information retrieval," *Annual Review of Information Science and Technology*, vol. 16, pp. 55–83, 1981.

[278] D. J. Urquhart, "The distribution and use of scientific and technical information," *The Royal Society Scientific Information Conference*, pp. 408–419, 1948.

[279] P. Vakkari, "Task-based information searching," *Annual Review of Information Science and Technology*, vol. 37, pp. 413–464, 2003.

[280] P. Vakkari, "Changes in search tactics and relevance judgments when preparing a research proposal: A summary of the findings of a longitudinal study," *Information Retrieval*, vol. 4, pp. 295–310, 2004.

[281] P. Vakkari and N. Hakala, "Changes in relevance criteria and problem stages in task performance," *Journal of Documentation*, vol. 56, pp. 540–562, 2000.

[282] P. Vakkari and K. Järvelin, "Explanation in information seeking and retrieval," in *New Directions in Cognitive Information Retrieval*, (A. Spink and C. Cole, eds.), pp. 113–138, Berlin: Springer, The Information Retrieval Series, 2005.

[283] P. Vakkari and E. Sormunen, "The influence of relevance levels on the effectiveness of interactive information retrieval," *Journal of the American Society for Information Science and Technology*, vol. 55, pp. 963–969, 2004.

[284] A. Veerasamy and N. J. Belkin, "Evaluation of a tool for visualization of information retrieval results," in *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 85–92, Zurich, Switzerland, 1996.

[285] A. Veerasamy and R. Heikes, "Effectiveness of a graphical display of retrieval results," *SIGIR Forum*, vol. 31, pp. 236–245, 1997.

[286] M. Viswanathan, *Measurement Error and Research Design*. Sage Publications, Inc, 2005.

[287] E. M. Voorhees, "On test collections for adaptive information retrieval," *Information Processing and Management*, vol. 44, pp. 1879–1885, 2008.

Full text available at: http://dx.doi.org/10.1561/1500000012

[288] E. M. Voorhees and D. K. Harman, *TREC: Experiment and Evaluation in Information Retrieval*. Cambridge, MA: MIT Press, 2005.

[289] J. B. Walther, "Research ethics in Internet-enabled research: Human subjects issues and methodological myopia," *Ethics and Information Technology*, vol. 4, pp. 205–216, 2002.

[290] P. Wang, "Methodologies and methods for user behavior research," *Annual Review of Information Science and Technology*, vol. 34, pp. 53–99, 1999.

[291] L. Wen, I. Ruthven, and P. Borlund, "The effects on topic familiarity on online search behaviour and use of relevance criteria," in *Proceedings of the 28th European Conference in Information Retrieval* (*ECIR 2006*), London, UK, 2006.

[292] R. W. White, M. Bilenko, and S. Cucerzan, "Studying the use of popular destinations to enhance web search interaction," in *Proceedings of the 30th Annual ACM International Conference on Research and Development in Information Retrieval* (*SIGIR '07*), pp. 159–166, Amsterdam, The Netherlands, 2007.

[293] R. W. White, I. Ruthven, J. M. Jose, and C. J. van Rijsbergen, "Evaluating implicit feedback models using searcher simulations," *ACM Transactions on Information Systems*, vol. 23, pp. 325–261, 2005.

[294] B. M. Wildemuth, *Applications of Social Research Methods to Questions in Information and Library Science*. Libraries Unlimited, (in press).

[295] B. M. Wildemuth, M. Yang, A. Hughes, R. Gruss, G. Geisler, and G. Marchionini, "Access via features versus access via transcripts: User performance and satisfaction," *TREC VID 2003 Notebook Paper*, 2003.

[296] F. Williams, *Reasoning with Statistics: How to Read Quantitative Research*. Orlando, Florida: Holt, Rinehart and Winston, Inc, Fourth ed., 1992.

[297] T. D. Wilson, "On user studies and information needs," *Journal of Documentation*, vol. 37, pp. 3–15, 1981.

[298] L. Xiong and E. Agichtein, "Towards privacy-preserving query log publishing," in *Proceedings of the Workshop on Query Log Analysis: Social and Technology Challenges at the 16th International World Wide Web Conference*, Banff, Canada, 2007.

[299] W. Yuan and C. T. Meadow, "A study of the use of variables in information retrieval user studies," *Journal of the American Society for Information Science*, vol. 50, pp. 140–150, 1999.

[300] P. Zhang, L. Plettenberg, J. L. Klavans, D. W. Oard, and D. Soergel, "Task-based interaction with an integrated multilingual, multimedia information system: A formative evaluation," in *Proceedings of the 7th ACM/IEEE Joint Conference on Digital Libraries* (*JCDL '07*), pp. 117–126, Vancouver, BC, 2007.

[301] X. Zhang, "Collaborative relevance judgments: A group consensus method for evaluating user search performance," *Journal of the American Society for Information Science and Technology*, vol. 53, pp. 220–231, 2002.