
Mark D. Smucker

Department of Management Sciences
University of Waterloo
200 University Avenue West
Waterloo, Ontario, Canada N2L 3G1

mark.smucker@uwaterloo.ca
<http://www.mansci.uwaterloo.ca/~msmucker/>
Phone: +1 (519) 888-4567 ext. 38620
Fax: +1 (519) 746-7252

Education

University of Massachusetts Amherst

- Ph.D. in Computer Science, September 2008.

University of Wisconsin-Madison

- M.S. in Computer Science, May 1996.

Iowa State University

- B.S. in Computer Science with Distinction, May 1994.
- B.S. in Physics with Distinction, May 1994.

Research Interests

I am interested in the design, analysis, and evaluation of interactive information retrieval systems.

Awards

- ACM SIGIR 2012 Best Paper Award.
- Faculty of Engineering Teaching Excellence Award, University of Waterloo, 2012.
- Outstanding Performance Award (OPA), University of Waterloo, May 2013.
- University of Waterloo Engineering Society Teaching Excellence Honourable Mention, 2014.
- Faculty of Engineering Distinguished Performance Award, University of Waterloo, July 2016.

Student Honors

- Finalist (reached last round of competition), Microsoft Research Graduate Fellowship, 2007.
- Honor Societies: Phi Beta Kappa, Phi Kappa Phi, and Upsilon Pi Epsilon.
- Earned 3 undergraduate scholarships and 2 departmental awards for academic performance.
- Iowa State University Scholarship of Distinction: 20 awarded to incoming freshmen, full year of tuition, fall 1989.

Research Experience

Associate Professor, July 2014 to present. (Assistant Professor, July 2008 to June 2014).

Department of Management Sciences, University of Waterloo.

- Introduced time-biased gain (TBG) with Charles Clarke [9]¹ (SIGIR Best Paper). TBG produces validated predictions of human performance for information retrieval systems.
- Created a stochastic simulation of time-biased gain that estimates a user population's distribution of gain on a per-topic basis. With these distributions we can measure the effect-size of improvements and better understand the value of the improvements [7, 28].
- Demonstrated that retrieval precision does significantly affect user performance, and that the better the match between retrieval metric and user interface, the better the metric predicts user performance [10].
- With collaborators, showed the impact of spam on web retrieval and developed an efficient and effective means to detect and remove spam to improve retrieval quality [1].

Visiting Researcher, Microsoft Research, Redmond, WA. May 2015 to August 2015.

- Modeled user behavior on the Bing search results page in collaboration with Ryen White, Sue Dumais, and Paul Bennet.

¹ All citation numbers refer to the list of numbered publications that starts on page 3.

Graduate Research Assistant, Center for Intelligent Information Retrieval, University of Massachusetts Amherst. Advisor: James Allan. August 2003 to June 2008.

- Studied find-similar (“more like this” or “related articles”) and found it to have the potential to match the performance of multiple item relevance feedback [14].
- Studied statistical significance tests for information retrieval evaluation and showed that researchers should be careful in their use of the Wilcoxon and sign tests [12].

Undergraduate, Iowa State University, Ames, Iowa, 1993-1995, and Graduate Research Assistant, Los Alamos National Lab’s Center for Nonlinear Studies, Los Alamos, New Mexico, summer 1994.

- Discovered and analyzed social networks in a system that modeled social choice and refusal [2, 16, 68]. Studied the effect of population structures on genetic algorithms [15].

Teaching Experience

Department of Management Sciences, University of Waterloo. July 2008 to present.

- Developed and taught 4 new courses: Introduction to Computer Programming (GENE/MSCI 121), Data Structures and Algorithms (MSCI 240), Data Mining (MSCI 446), and Information Retrieval (MSCI 541 / 720).
- In 2015, revised Principles of Software Engineering (MSCI 342) to focus on Agile software development as a team.
- Completed 4-day Teaching Excellence Academy (TEA) in 2015, and completed 3-day Instructional Skills Workshop (ISW) in 2009.

Department of Computer Science, University of Massachusetts Amherst. Spring 2007.

- Fully responsible for all aspects of a 1-credit, 8-week long, C++ programming course.

Department of Computer Sciences, University of Wisconsin-Madison. Fall 1994 to spring 1996.

- Taught C++ programming: two sections fall 1995-spring 1996, and one section, spring 1995.
- Earned an average score of 4.6 / 5 for “would recommend instructor” on student evaluations spring 1996. Only 2 of the other 14 instructors received higher ratings.
- Worked as computer lab consultant supporting students, fall 1994 and summer 1995.

Industry Experience

Software Contractor, Seattle, WA. May 2002 to July 2003.

- Wrote specifications and worked directly with customers to determine requirements. Wrote custom software related to e-mail programs for Jetstream Software.

Senior Research Analyst, Avenue A, Inc., Seattle, WA. September 1999 to April 2002.

- Performed data analyses across advertisers and answered specific client questions.
- Originated idea for a distributed cluster to allow high speed, scalable processing of terabytes worth of click-stream data. This architecture became a core part of Avenue A’s service offering.

Software Design Engineer, Microsoft, Redmond, WA. April 1998 to September 1999.

- Worked as both engineer and program manager in an advanced technology team creating software that adapted to user behavior.

Engineer, Firefly (aka Agents, Inc.), Cambridge, MA. June 1996 to April 1998 (acquired by Microsoft).

- Wrote front end for a part of My Yahoo that provided website recommendations via collaborative filtering. Performed analyses of server load and evaluation of the quality of recommendations.

Professional Licence

- Limited Engineer Licence (LEL) in the area of Software, Professional Engineers Ontario. 2011-present.

Publications

Journal Articles

1. Gordon V. Cormack, Mark D. Smucker, and Charles L. A. Clarke. "Efficient and Effective Spam Filtering and Re-Ranking for Large Web Datasets," *Journal of Information Retrieval*, Volume 14, Issue 5, 2011, pp. 441-465.
2. Dan Ashlock, Mark D. Smucker, E. Ann Stanley, and Leigh Tesfatsion. "Preferential Partner Selection in an Evolutionary Study of Prisoner's Dilemma," *BioSystems*, 1996, Vol. 37, No 1-2, pp. 99-125.

Refereed Conference Papers (Full papers, 8+ pages in length)

3. Gaurav M. Baruah, Mark D. Smucker, and Charles L. A. Clarke, "Evaluating Streams of Evolving News Events," In the Proceedings of the 38th Annual ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, 2015, ACM Press, pages 675-684. (Conference acceptance rate: $70/351 = 20\%$)
4. Charles L. A. Clarke and Mark D. Smucker, "Time Well Spent," In the Proceedings of the Fifth Information Interaction in Context Symposium (IIIX'14), Regensburg, Germany, 2014, ACM Press, pages 205-214. (Conference acceptance rate: $21/45 = 46\%$)
5. Aiman L. Al-Harbi and Mark D. Smucker, "A Qualitative Exploration of Secondary Assessor Relevance Judging Behavior," In the Proceedings of the Fifth Information Interaction in Context Symposium (IIIX'14), Regensburg, Germany, 2014, ACM Press, pages 195-204. (Conference acceptance rate: $21/45 = 46\%$)
6. Le Li and Mark D. Smucker. "Tolerance of Effectiveness Measures to Relevance Judging Errors," In the Proceedings of the European Conference on Information Retrieval (ECIR 2014), Springer International Publishing, LNCS 8416, pages 148-159, 2014. (Conference acceptance rate: $33/143 = 23\%$)
7. Mark D. Smucker and Charles L. A. Clarke. "Modeling User Variance in Time-Biased Gain," In the Proceedings of the Sixth Symposium on Human-Computer Interaction and Information Retrieval (HCIR'12), 2012, ACM Press. 10 pages. (Conference acceptance rate: $4/11 = 36\%$)
8. Mark D. Smucker, James Allan, and Blagovest Dachev. "Human Question Answering Performance Using an Interactive Document Retrieval System," In the Proceedings of the Fourth Information Interaction in Context Symposium (IIIX'12), Nijmegen, The Netherlands, 2012, ACM Press, pp. 35-44. (Conference acceptance rate: $25/40 = 63\%$)
9. Mark D. Smucker and Charles L. A. Clarke. "Time-Based Calibration of Effectiveness Measures," In the Proceedings of the 35th Annual ACM SIGIR Conference on Research and Development in Information Retrieval, Portland, OR, USA, 2012, ACM Press, pp. 95-104. (This paper won the **Best Paper Award**. Award rate: $1/483 = 0.2\%$. Conference acceptance rate: $98/483 = 20\%$.)
10. Mark D. Smucker and Chandra Prakash Jethani. "Human Performance and Retrieval Precision Revisited," In Proceedings of SIGIR'10: 33rd Annual ACM SIGIR Conference on Research and Development in Information Retrieval, Geneva, Switzerland, 2010, ACM Press, pp. 595-602. (Conference acceptance rate: $87/520 = 17\%$)
11. Jimmy Lin and Mark D. Smucker. "How Do Users Find Things with PubMed? Towards Automatic Utility Evaluation with User Simulations," In the Proceedings of the 31st Annual

ACM SIGIR Conference on Research and Development in Information Retrieval, Singapore, 2008, ACM Press, pp. 19-26. (Conference acceptance rate: $85/496 = 17\%$)

12. Mark D. Smucker, James Allan, and Ben Carterette. "A Comparison of Statistical Significance Tests for Information Retrieval Evaluation," In the Proceedings of CIKM'07: 16th ACM International Conference on Information and Knowledge Management, Lisboa, Portugal, 2007, ACM Press, pp. 623-632. (Conference acceptance rate: $87/512 = 17\%$)
13. Ben Carterette and Mark D. Smucker. "Hypothesis Testing with Incomplete Relevance Judgments," In the Proceedings of CIKM'07: 16th ACM International Conference on Information and Knowledge Management, Lisboa, Portugal, 2007, ACM Press, pp. 643-652. (Conference acceptance rate: $87/512 = 17\%$)
14. Mark D. Smucker and James Allan. "Find-Similar: Similarity Browsing as a Search Tool," In the Proceedings of the 29th Annual international ACM SIGIR Conference on Research and Development in Information Retrieval, 2006, ACM Press, pp. 461-468. (Conference acceptance rate: $74/399 = 19\%$)
15. Dan Ashlock, Mark D. Smucker, and John Walker. "Graph Based Genetic Algorithms," In the Proceedings of the 1999 Congress on Evolutionary Computation, 1999, pp. 1362-1368.
16. E. Ann Stanley, Dan Ashlock, and Mark D. Smucker. "Iterated Prisoner's Dilemma with Choice and Refusal of Partners: Evolutionary Results," In the Proceedings of the Third European Conference on Advances in Artificial Life, Lecture Notes In Computer Science, vol. 929, Springer-Verlag, 1995, pp. 490-502.

Book Chapters

17. Mark D. Smucker. "Information Representation," Chapter 5 of *Interactive Information Seeking, Behaviour and Retrieval*, Editors: Ian Ruthven and Diane Kelly, Facet Publishing, 2011, pp. 77-94. (Invited by Editors)

Refereed Short Conference Papers (4 pages or longer)

18. Gaurav Baruah, Haotian Zhang, Rakesh Guttikonda, Jimmy Lin, Mark D. Smucker, and Olga Vechtomova. "Optimizing Nugget Annotations with Active Learning," In the Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (CIKM), 2016, pp. 2359-2364. (Conference acceptance rate: 935 submissions, 165 accepted as long paper, and 105 accepted as short papers = 28.9%).
19. Haotian Zhang, Jimmy Lin, Gordon V. Cormack, and Mark D. Smucker. "Sampling Strategies and Active Learning for Volume Estimation," In the Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, 2016, pp. 981-984. (Conference acceptance rate: $104/339 = 30.7\%$)
20. Aiman Al-Harbi and Mark D. Smucker. "Are Secondary Assessors Uncertain When They Disagree About Relevance Judgements?" In the Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval (CHIIR), 2016, pp. 233-236. (Conference acceptance rate: $27/54 = 50\%$)

21. Mark D. Smucker and Charles L. A. Clarke. "Modeling Optimal Switching Behavior," In the Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval (CHIRR) 2016, pp. 317-320. (Conference acceptance rate: $27/54 = 50\%$)
22. Luchen Tan, Haotian Zhang, Charles L. A. Clarke, and Mark D. Smucker. "Lexical comparison between Wikipedia and twitter corpora by using word embeddings," In the Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Beijing, China, 2015, pages 657-661. (Conference acceptance rate: 22%).
23. Gaurav Baruah, Adam Roegiest, and Mark D. Smucker. "Pooling for User-Oriented Evaluation Measures," In the Proceedings of the 2015 International Conference on The Theory of Information Retrieval (ICTIR), ACM, 2015, pages 341-344. (Conference acceptance rate: $22/43 = 51\%$)
24. Mark D. Smucker and Charles L. A. Clarke, "Modeling Optimal Switching Behavior," In the Proceedings of the First ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR), Chapel Hill, North Carolina, USA, 2016, ACM Press, 4 pages. (Conference acceptance rate: $27/54 = 50\%$)
25. Aiman L. Al-Harbi and Mark D. Smucker, "Are Secondary Assessors Uncertain When They Disagree About Relevance Judgements?," In the Proceedings of the First ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR), Chapel Hill, North Carolina, USA, 2016, ACM Press, 4 pages. (Conference acceptance rate: $27/54 = 50\%$)
26. Gaurav M. Baruah, Adam Roegiest, and Mark D. Smucker, "The Effect of Expanding Relevance Judgements with Duplicates," In the Proceedings of the 37th Annual ACM SIGIR Conference on Research and Development in Information Retrieval, Gold Coast, Queensland, Australia, 2014, ACM Press, pp. 1159-1162. (Conference acceptance rate: $104/263 = 40\%$)
27. Mark D. Smucker, Xiaoyu Sunny Guo, and Andrew Toulis, "Mouse Movement During Relevance Judging: Implications for Determining User Attention," In the Proceedings of the 37th Annual ACM SIGIR Conference on Research and Development in Information Retrieval, Gold Coast, Queensland, Australia, 2014, ACM Press, pp. 978-982. (Conference acceptance rate: $104/263 = 40\%$)
28. Mark D. Smucker and Charles L. A. Clarke. "Stochastic Simulation of Time-Biased Gain," In the Proceedings of CIKM, 2012, ACM Press, pp. 2040-2044. (Accepted as a short paper. 1088 submissions. 13.4% accepted as regular papers and 14.3% as short papers. Overall acceptance rate: 28%.)
29. Mark D. Smucker and James Allan. "A New Measure of the Cluster Hypothesis," In the Proceedings of the 2nd International Conference on the Theory of Information Retrieval (ICTIR'09), September 2009, Cambridge, UK, pp. 281-288. (Accepted as a short paper. 82 submissions. 22% accepted as regular papers. 21% accepted as short papers. Overall paper acceptance rate = 43%.)

Refereed Poster (Short 2-Page) Conference Papers

30. Mark D. Smucker and Chandra Prakash Jethani. "Time to Judge Relevance as an Indicator of Assessor Error," In the Proceedings of the 35th Annual ACM SIGIR Conference on Research and Development in Information Retrieval, Portland, OR, USA, 2012, ACM Press, pp. 1153-1154. (Conference acceptance rate: $76/208 = 37\%$)

31. Mark D. Smucker and Chandra Prakash Jethani. "Measuring Assessor Accuracy: A Comparison of NIST Assessors and User Study Participants," In the Proceedings of the 34th Annual ACM SIGIR Conference on Research and Development in Information Retrieval, Beijing, China, July 2011, ACM Press, pp. 1231-1232. (Conference acceptance rate: $89/274 = 33\%$)
32. Mark D. Smucker, James Allan, and Ben Carterette. "Agreement Among Statistical Significance Tests for Information Retrieval Evaluation at Varying Sample Sizes," In Proceedings of SIGIR'09: 32nd Annual ACM SIGIR Conference on Research and Development in Information Retrieval, Boston, Massachusetts, USA, 2009, ACM Press, pp. 630-631. (Conference acceptance rate: $86/256 = 34\%$)
33. Mark D. Smucker and James Allan. "Using Similarity Links as Shortcuts to Relevant Web Pages," In Proceedings of the 30th Annual ACM SIGIR Conference on Research and Development in Information Retrieval, 2007, ACM Press, pp. 863-864. (Conference acceptance rate: $105/235 = 45\%$)
34. Mark D. Smucker and James Allan. "Lightening the Load of Document Smoothing for Better Language Modeling Retrieval," In Proceedings of the 29th Annual international ACM SIGIR Conference on Research and Development in Information Retrieval, 2006, ACM Press, pp. 699-700. (Conference acceptance rate: $52/140 = 37\%$)

Refereed Workshop Papers

35. Aiman L. Al-Harbi and Mark D. Smucker. "User Expressions of Relevance Judgment Certainty," In the Proceedings of the 7th Annual Symposium on Human-Computer Interaction and Information Retrieval (HCIR 2013), October 2013, Vancouver, BC, Canada. 4 pages. (Workshop short paper acceptance rate: $16/16 = 100\%$)
36. Mark D. Smucker and Charles L. A. Clarke. "The fault, dear researchers, is not in Cranfield, But in our metrics, that they are unrealistic," In the Proceedings of EuroHCIR, August 2012, Nijmegen, The Netherlands. 2 pages. (This paper was 1 of 9 papers selected for oral presentation. Workshop acceptance $9/32 = 28\%$.)
37. Mark D. Smucker. "An Analysis of User Strategies for Examining and Processing Ranked Lists of Documents," In the Proceedings of the Fifth Workshop on Human-Computer Interaction and Information Retrieval (HCIR'11), October 20, 2011, Mountain View, CA, USA. 4 pages. (Accepted for a short oral presentation.)
38. Mark D. Smucker and Chandra Prakash Jethani. "The Crowd vs. the Lab: A Comparison of Crowd-Sourced and University Laboratory Participant Behavior," In the Proceedings of the SIGIR 2011 Workshop on Crowdsourcing for Information Retrieval, July 2011, Beijing, China. 6 pages. (This paper won the workshop's Crowdsourcing Challenge Contest.)
39. Mark D. Smucker and Chandra Prakash Jethani. "Impact of Retrieval Precision on Perceived Difficulty and Other User Measures," In the Proceedings of the Fourth Workshop on Human-Computer Interaction and Information Retrieval (HCIR'10), August 22, 2010, New Brunswick, NJ, USA. 4 pages. (This paper was 1 of 6 papers accepted for oral presentation.)
40. Chandra Prakash Jethani and Mark D. Smucker. "Modeling the Time to Judge Document Relevance," In the Proceedings of the SIGIR'10 Workshop on the Simulation of Interaction, July 23, 2010, Geneva, Switzerland. 2 pages.

41. Mark D. Smucker. "Towards Timed Predictions of Human Performance for Interactive Information Retrieval Evaluation," In the Proceedings of the Third Workshop on Human-Computer Interaction and Information Retrieval (HCIR'09), October 23, 2009, Washington DC. 4 pages.
42. Mark D. Smucker. "A Plan for Making Information Retrieval Evaluation Synonymous with Human Performance Prediction," In the proceedings of the SIGIR'09 Workshop on the Future of Information Retrieval Evaluation, Boston, 2009. 2 pages. (Acceptance rate: 61% = 20/33)
43. Mark D. Smucker. "Work in Progress: Navigating Document Networks," HCIR'07 Workshop on Human-Computer Interaction and Information Retrieval, October 23, 2007, Cambridge, Massachusetts. 2 pages.
44. Mark D. Smucker and James Allan. "Measuring the Navigability of Document Networks," SIGIR'07 Web Information-Seeking and Interaction Workshop, Amsterdam, The Netherlands, 2007. 4 pages.

Refereed Posters

45. Rebecca Reznik-Zellen, Bob Stevens, Michael Thorn, Jeff Morse, Mark D. Smucker, James Allan, David Mimno, Andrew McCallum, and Mark Tuominen. "InterNano: e-Science for the Nanomanufacturing Community," In Proceedings of the 4th IEEE International Conference on e-Science, 2008. 2 pages.

Refereed Tutorials

46. Charles L. A. Clarke, Mark D. Smucker, and Emine Yilmaz. "IR Evaluation: Modeling User Behavior for Measuring Effectiveness," In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2015, pages 1117-1120.

Thesis

47. Mark D. Smucker. "Evaluation of Find-Similar with Simulation and Network Analysis," Ph.D. Thesis, University of Massachusetts Amherst, September 2008. 150 pages.

Unrefereed Papers

48. Haotian Zhang, Wu Lin, Yipeng Wang, Charles L.A. Clarke and Mark D. Smucker. "WaterlooClarke: TREC 2015 Total Recall Track," In the Online Proceedings of the 2015 Text Retrieval Conference (TREC 2015), NIST. 8 pages. (Selected for presentation at TREC 2015.)
49. Krisztian Balog, David Elsweiler, Evangelos Kanoulas, Liadh Kelly, and Mark D. Smucker. "Report on the CIKM workshop on living labs for information retrieval evaluation," SIGIR Forum 48, (1), June 2014, ACM Press, pp. 21-28.
50. Charles L. A. Clarke, Luanne Freund, Mark D. Smucker, and Emine Yilmaz. "Report on the SIGIR 2013 Workshop on Modeling User Behavior for Information Retrieval Evaluation (MUBE 2013)," SIGIR Forum, Volume 47 (2), December 2013, ACM Press, pp. 84-95.
51. Diane Kelly, Jaime Arguello and Robert Capra (Editors). "NSF Workshop on Task-based Information Search Systems," SIGIR Forum, Volume 47 (2), December 2013, ACM Press, pp. 116-127. (I was a participant in the workshop, and I am a co-author of the report.)

52. Robert Capra, Luanne Freund, Catherine L. Smith, Mark D. Smucker, Ryen W. White. "HCIR 2013: The Seventh International Symposium on Human-Computer Interaction and Information Retrieval," SIGIR Forum, Volume 47 (2), December 2013, ACM Press, pp. 33-40.
53. Charles L. A. Clarke, Luanne Freund, Mark D. Smucker, and Emine Yilmaz. "SIGIR 2013 Workshop on Modeling User Behavior for Information Retrieval Evaluation," In the Proceedings of the 36th Annual ACM SIGIR Conference on Research and Development in Information Retrieval, 2013, Dublin, Ireland. 1 page.
54. Mark D. Smucker, Gabriella Kazai, and Matthew Lease, "Overview of the TREC 2013 Crowdsourcing Track", TREC 2013, 6 pages.
55. P. Robert Duimering, Samir Elhedhli, Beth Jewkes, and Mark D. Smucker. "Management Engineering: The Engineering of Management Systems," Report prepared for Engineers Canada, Department of Management Sciences, University of Waterloo, January 2013. 21 pages.
56. Mark D. Smucker, Gabriella Kazai, and Matthew Lease. "Overview of the TREC 2012 Crowdsourcing Track," In the Online Proceedings of the 2012 Text Retrieval Conference (TREC 2012), NIST. 9 pages.
57. Mark D. Smucker. "Crowdsourcing with a Crowd of One and Other TREC 2011 Crowdsourcing and Web Track Experiments," In the Online Proceedings of the 2011 Text Retrieval Conference (TREC 2011), NIST. 4 pages. (Selected for presentation at TREC 2011.)
58. Mark D. Smucker, Charles L. A. Clarke, Gordon V. Cormack, and Olga Vechtomova. "University of Waterloo at TREC 2010: Legal Interactive," In the Online Proceedings of the 2010 Text Retrieval Conference (TREC 2010), NIST. 3 Pages. (Selected for presentation at TREC 2010.)
59. Chris Buckley, Matthew Lease, Mark D. Smucker. "Overview of the TREC 2010 Relevance Feedback Track (Notebook)," In the Online Proceedings of the 2010 Text Retrieval Conference (TREC 2010), NIST. 4 pages.
60. Leif Azzopardi, Kalervo Järvelin, Jaap Kamps, and Mark D. Smucker. "Report on the SIGIR 2010 Workshop on the Simulation of Interaction," SIGIR Forum, Volume 44, December 2010, ACM Press, pp. 35-47.
61. Mark D. Smucker, Charles L. A. Clarke, and Gordon V. Cormack. "Experiments with ClueWeb09: Relevance Feedback and Web Tracks," In the Online Proceedings of the 2009 Text Retrieval Conference (TREC 2009), NIST. 5 pages.
62. Mark D. Smucker, James Allan, and Blagovest Dachev. "UMass Complex Interactive Question Answering (ciQA) 2007: Human Performance as Question Answerers," In the Online Proceedings of the 2007 Text Retrieval Conference (TREC 2007), NIST. 4 pages. (Selected for presentation at TREC 2007.)
63. Mark D. Smucker. "UMass Genomics 2006: Query-Biased Pseudo Relevance Feedback," In the Online Proceedings of the 2006 Text Retrieval Conference (TREC 2006), NIST. 4 pages.
64. Nasreen Abdul-Jaleel, James Allan, W. Bruce Croft, Fernando Diaz, Leah Larkey, Xiaoyan Li, Mark D. Smucker, and Courtney Wade. "UMass at TREC 2004: Novelty and HARD," In the Online Proceedings of the 2004 Text Retrieval Conference (TREC 2004), NIST. 13 pages.

Technical Reports (excludes technical reports that duplicate above published work)

65. Mark D. Smucker and James Allan. "An Investigation of Dirichlet Prior Smoothing's Performance Advantage," University of Massachusetts Amherst, Center for Intelligent Information Retrieval (CIIR), Technical Report IR-391, 2005. 8 pages.
66. Mark D. Smucker, Fernando Diaz, and James Allan. "High Precision Retrieval via User Interaction and Metadata," University of Massachusetts Amherst, Center for Intelligent Information Retrieval (CIIR), Technical Report IR-390, 2005. 8 pages.
67. Mark D. Smucker, David Kulp, and James Allan. "Dirichlet Mixtures for Query Estimation in Information Retrieval," University of Massachusetts Amherst, Center for Intelligent Information Retrieval (CIIR), Technical Report IR-445, April 2005. 13 pages.
68. Mark D. Smucker, E. Ann Stanley, and Dan Ashlock. "Analyzing Social Network Structures in the Iterated Prisoner's Dilemma with Choice and Refusal," University of Wisconsin-Madison Computer Sciences Tech Report CS-TR-94-1259, December 1994, revised August 1995. 37 pages.

Proceedings (Ed.)

69. Krisztian Balog, David Elsweiler, Evangelos Kanoulas, Liadh Kelly, and Mark D. Smucker (editors). Proceedings of the 2013 Workshop on Living Labs for Information Retrieval Evaluation. CIKM 2013, San Francisco, 2013, ACM.
70. Charles L. A. Clarke, Luanne Freund, Mark D. Smucker, and Emine Yilmaz (editors). Proceedings of the SIGIR 2013 Workshop on Modeling User Behavior for Information Retrieval Evaluation, Dublin, Ireland, August 1, 2013.
71. Robert Capra, Luanne Freund, Gene Golovchinsky, Catherine Smith, Mark D. Smucker, Ryan White (editors). The Online Proceedings of the 7th Annual Symposium on Human-Computer Interaction and Information Retrieval (HCIR 2013), Vancouver, 2013.
72. Leif Azzopardi, Kalervo Järvelin, Jaap Kamps, and Mark D. Smucker (editors). Proceedings of the SIGIR 2010 Workshop on the Simulation of Interaction: Automated Evaluation of Interactive IR, Geneva, Switzerland, July 23, 2010.

Articles in Preparation

73. Mark D. Smucker, Gaurav Baruah, Le Li, Gabriella Kazai, and Matthew Lease. "Evaluating Crowdsourced Relevance Judgments," to be submitted to a journal.
74. Chandra Prakash Jethani and Mark D. Smucker. "Effect of Prevalence on Relevance Assessing Behaviour," thesis to be revised and submitted to a journal.
75. Mark D. Smucker. "An Analysis of User Behavior when Examining and Processing Ranked Lists of Documents," an extended version of my HCIR 2011 workshop paper to be submitted to a journal.

Presentations (Excluding Above Publications)

1. Invited member of "pro" team for debate on "Simulating Users is the Way Forward in Evaluation of Interactive Information Retrieval," Information Interaction in Context Symposium (IIiX), New Brunswick, NJ, USA, 2010. Other team members: Leif Azzopardi, Kalervo Järvelin, and Jaap Kamps.

2. “The Wonderful Category B of ClueWeb09,” TREC 2009 poster. Co-author: Gordon V. Cormack.

Research Funding

Year	Amount	Source
2015-2017	\$179,691 (\approx \$60,000 per year)	Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research and Development grant (CRD). Industry partner: Google. Co-PIs: Charles L. A. Clarke and Gordon V. Cormack.
2014-2018	\$32,000 per year, total: \$160,000	Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant
2015	\$71,500 (in-kind value)	2015 Compute Canada Resource Allocation (RAC) for 65 TB of storage.
2014	\$10,778	GRAND (Graphics, Animation and New Media) NCE
2009-2013	\$24,000 per year, total: \$120,000	Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant
2012	\$5000	NSERC Supplemental Discovery Grant
2009-2013	US\$10,000 (Apr. 2013) US\$7500 (Mar. 2011) US\$3000 (Dec. 2009) US\$4000 (June 2009) total: US\$24,500	Amazon Web Services in Education Research Grant Dec. 2009 grant with Charles L. A. Clarke and Gordon V. Cormack; my share of the award was 100%.
2011	\$200	SIGIR 2011 CrowdSourcing Workshop Challenge Grant, CrowdFlower, Inc.

Teaching

Courses Taught in the Past 5 Years

In addition to teaching each of these courses, I developed each of these courses for the Management Engineering undergraduate program. The average class size, as of fall 2013, is listed as “Avg. Size”.

Title	Level	Avg. Size
MSCI 121 (was GENE) Introduction to Computer Programming	1B (Core)	67.5
MSCI 240 Data Structures and Algorithms	2A (Core)	50.6
MSCI 342 Principles of Software Engineering	3B (Core)	46
MSCI 446 Data Mining	3A+ (Tech. Elective)	28.0
MSCI 541/720 Information Retrieval	3A+ (Tech. Elective) and Graduate	14.5

Course Critique Scores

Term and year	Course	Number of Students	Critique Q 10	Critique Avg. Q 1-9	% Response
Fall 2016	MSCI 342	52	77	82	79%
Winter 2016	MSCI 541/720	15	73	81	87%
Fall 2015	MSCI 342	46	78	83	76%
Fall 2014	MSCI 240	58	91	91	78%
Winter 2014	MSCI 541/720	14	87	90	79%
Winter 2014	MSCI 121	74	86	87	76%
Fall 2013	MSCI 240	51	92	92	76%
Winter 2013	MSCI 541/720	18	92	90	56%
Winter 2013	GENE 121	60	89	91	67%
Fall 2012	MSCI 240	61	92	89	67%
Winter 2012	MSCI 541/720	11	94	93	82%
Winter 2012	GENE 121	64	81	86	84%
Winter 2011	GENE 121	74	78	81	53%
Fall 2010	MSCI 446	28	81	81	86%
Fall 2010	MSCI 240	54	76	78	67%
Winter 2010	GENE 121	72	76	76	61%
Fall 2009	MSCI 240	46	69	75	76%
Fall 2008	MSCI 240	41	89	86	71%
Averages:		46.6	83.4	85.1	73.3%

Note: Question 10 is “Quality of Teaching,” and Questions 1-9 address attributes of teacher quality.

Student Supervision

Current Graduate Students

Name (Co-Supervisor)	Degree	Supervisory Period	Research Description and Comments
Haotian Zhang (Gordon Cormack)	PhD	1/2015-	Haotian is searching for a thesis topic.
Mustafa Abualsaud	MMath	1/2015-	Mustafa is studying how search systems can

	Thesis		aid learning new material.
Shahin Rahbariasl	MMath Thesis	9/2016-	Shahin is investigating the effect of judging speed on relevance judging.
Nimesh Ghelani	MMath Thesis	9/2016-	Nimesh is new and is searching for a thesis topic.

Former Graduate Students

Name (Co-Supervisor)	Degree	Supervisory Period	Research Title and Comments
Aiman L. Al-Harbi	PhD	3/2011- 8/2016	<i>Studying Relevance Judging Behavior of Secondary Assessors</i> Post-degree: Assistant Prof., King Saud bin Abdulaziz University for Health Sciences
Gaurav Baruah (Charles Clarke)	PhD	9/2011- 8/2016	<i>Filtering News from Document Streams: Evaluation Aspects and Modeled Stream Utility</i> Post-degree: Post Doctoral Fellow, University of Waterloo
Le Li	MMath Thesis	9/2011- 8/2013	<i>Evaluating Information Retrieval Systems With Multiple Non-Expert Assessors</i> Post-degree: Software Engineer in Data, Electronic Arts, Inc., USA.
Chandra Prakash Jethani	MMath Thesis	9/2009- 8/2011	<i>Effect of Prevalence on Relevance Assessing Behaviour</i> Post-degree: Software Development Engineer, Yahoo! Inc., USA.
Xiaoyin (Susan) Yang (P. Robert Duimering)	MASc	1/2009- 10/2009	<i>Collective Intelligence in Collaborative Tagging System</i>

Current and Former Undergraduate Students

Name	Work Supervised	Supervisory Period	Research Title and Comments
Madeline Cowie, David Erakovic, Mojan Hamed, Matthew Watson	4th year MGTE design project	Spring 2016- Winter 2017	I am regularly meeting and advising this MGTE 4A/4B students on their project.
Frances Pogacar	Research Assistant	Fall 2014, Spring 2015	In the fall, Frances (4A CS) met with me to design her study that will investigate the effect of poor search results on user behavior. In the spring, Frances began the implementation of the study.

Hannah Gautreau	URA	Spring 2014	Hannah (2B MGTE) worked with me on the early stages of an experiment to investigate the query reformulation process.
Kevin Choi	URI	Spring 2014	Kevin (1B MGTE) wrote software in support of my research program.
Evelyn Tchao, Casey Tjokroharjo, Ron Yu, Amy Zhang	4th year MGTE design project	Spring 2014-Winter 2015	Regularly met and advised this MGTE 4A/4B students on their neighborhood recommendation system.
Xiaoyu (Sunny) Guo and Andrew Toulis	URA	Fall 2013	Sunny (2A MGTE) and Andy (2A MGTE) studied patterns of mouse movement during relevance assessing.
Hassan Ahmed	URA	Fall 2013	Hassan (2A MGTE) is studying distributions of time spent searching.
Eugene Solodkin	Research Assistant	Spring 2013	Eugene (4B SE) investigated methods for near duplicate detection of web pages. Post-degree: Software Development Engineer, Amazon, Inc.
Lamin Ceesay, Jonathan Ong, Jiansen Wang, Chad Xu	4th year MGTE design project	Spring 2013-Winter 2014	Regularly met and advised these MGTE 4A/4B students on their project titled <i>Optimizing Franchise Site Expansion</i> .
Jared Evans and Brian (Jea Hwan) Kim	URA	Spring 2013	Jared (4A MGTE) and Brian (2B MGTE) worked with me to study reading speed while people search for relevant information.
Joe Mercer	URA	Fall 2012	Joe (2A MGTE) performed a qualitative study about user needs for music recommendations.
Ziyad Mir	URA	Spring 2012	Ziyad (3A SYDE) and I looked at developing an experimental methodology to collect speed-accuracy trade-off data for relevance judgments.
Korey Cheung, Matthew Guilherme, Dairai Nyabando, and Alphonse Vermeulen	4th year MGTE design project	Spring 2011-Winter 2012	Regularly met and advised these MGTE 4A/4B students on their project titled <i>Investment Visualization Tool</i> .
Melissa Deziel	URA	Fall 2011	Melissa (3B MGTE) worked with me on models of user behavior and helped with initial investigations of the application of the diffusion decision model to problems in information retrieval. Post-degree: Product Associate, MediaMath

Kenneth (Chi Wai) Ng	URA	Fall 2011	Ken (2B SYDE) investigated the use of the Hadoop framework on SHARCNET.
David Hu	URA	Spring 2011	David (2B SE) implemented Broder's near duplicate detection algorithm on Amazon's EC2 Map Reduce system to allow for the study of the effect of near duplicates on information retrieval performance.
Mack Duan	URA	Spring 2011	Mack (2B SE) built a web-based front end for two retrieval systems to support a future user study on interactive web retrieval.
Pativet Sathiamsamrit and Joonha Shin	CBET Enterprise Co-op	Winter 2011	I mentored Pativet and Joonha (both 3B MGTE) during their entrepreneurial co-op.
Kevin Burt	URI/Co-op	Spring 2010	Kevin, Michael, and Chad (all 1B MGTE) assisted me with my research. Each worked full time (35 hr. / week) and I met with them daily. Projects centered around the analysis of user behavior in information retrieval systems.
Michael Tatham	URI/Co-op	Spring 2010	
Chad Xu	Volunteer Undergrad. Research Asst.	Spring 2010	
Johnathan Fisseha	URA	Spring 2009	Johnathan worked with me on a project about predicting the adoption of new user interaction mechanisms in search interfaces. Johnathan is a Management Engineering student.
Ching-Pei (Ben) Lin	URA, Independent Study	Spring 2009, Winter 2010	Ben and Ivan worked as a pair in S09. We read and discussed research papers together. Ben and Ivan contributed to my research for TREC 2009 (a yearly conference in my field). Ben came back, on his own, in W10 for independent study. Post-degree: Ben went to graduate school at UT-Austin in ECE, and Ivan when to graduate school at UW in ECE.
Ivan Patarroyo	URA	Spring 2009	
Jiao Yung (Justin) Lin	URA	Spring 2009	Justin's work centered around learning the advanced data processing technique known as Map Reduce and creating a prototype of its use with the Hadoop processing environment. Post-degree: Software Engineer, Google.

Thesis Examinations (PhD Committee or Master's Thesis Reader)*Graduated Students*

Year	Name	Degree	University
2016	Luchen Tan	PhD	University of Waterloo (Computer Science)
2015	Yuheng (Helen) Jiang	MASc	University of Waterloo (Management Sciences)
2014	Maryam Bashir	PhD	Northeastern University (Computer Science)
2014	Andrew Kane	PhD	University of Waterloo (Computer Science)
2014	Feza Baskaya	PhD	University of Tampre (Information Science)
2014	Adriel Dean-Hall	MMath	University of Waterloo (Computer Science)
2013	Mehrnaz Moustafapour	MASc	University of Waterloo (Management Sciences)
2013	Gobaan Raveendran	MMath	University of Waterloo (Computer Science)
2013	Shahab Kamali	PhD	University of Waterloo (Computer Science)
2013	Younes Abouelnagah	MMath	University of Waterloo (Computer Science)
2013	Azin Ashkan	PhD	University of Waterloo (Computer Science)
2013	Maheedhar Kolla	PhD	University of Waterloo (Computer Science)
2012	John A. Akinyemi	PhD	University of Waterloo (Computer Science)
2012	Mona Mojdeh	PhD	University of Waterloo (Computer Science)
2012	Mohammed Alliheedi	MMath	University of Waterloo (Computer Science)
2012	Adam Roegiest	MMath	University of Waterloo (Computer Science)
2011	Hani K. Nikkhoo	MMath	University of Waterloo (Computer Science)
2010	Kelly Y. Itakura	PhD	University of Waterloo (Computer Science)
2008	Rana Alshaar	MASc	University of Waterloo (Management Sciences)
2008	David R. A. Henriques	MASc	University of Waterloo (Management Sciences)
2008	Kun Cen	MASc	University of Waterloo (Management Sciences)

Current Students (Thesis Proposal or Agreed to be Master's Thesis/Project Reader)

Year	Name	Degree	University
2017	Adam Roegiest	PhD	University of Waterloo (Computer Science)
2016	Shaikh Quader	MMath	University of Waterloo (Computer Science)
2016	Ayman Alzayat	PhD	University of Waterloo (Management Sciences)
2013	Kourosh Malek	PhD	University of Waterloo (Management Sciences)

Service**Committees and Other Roles***Department of Management Sciences*

Period	Name and Role
Fall 2015 to present	Associate Chair, Undergraduate Studies
Spring 2009 to Spring 2016	Examinations & Promotions Officer
Fall 2012 to present	Undergraduate Studies Committee (DUGS) - Member
Fall 2012 and Fall 2010	Management Engineering 2A Class professor
Sept. 2010 to June 2011 Nov. 2009 to April 2010	Department Advisory Committee on Appointments (DACA) - Member
Winter 2009 to Fall 2010	Computing Representative – Facilitate computing solutions in department

I have also served the department with various ad-hoc service. Selected examples:

- Faculty representative at Fall Open House, 2014, 2015, 2016.
- Faculty representative at Ontario University Fair, Fall 2015, 2016.
- Co-author of opinion paper on Management Engineering program for Engineers Canada, 2012.
- Faculty participant in Co-op stakeholder feedback meeting, May 7, 2012.
- Faculty representative at Engineering Student Awards dinner, Nov. 11, 2010 and Nov. 12, 2009.
- March Break meet and greet prospective students, March 17, 2009.

Faculty of Engineering

Period	Name and Role
Spring 2009 to Spring 2016	Examinations & Promotions Committee – Member
Fall 2015 to present	Faculty Undergraduate Studies Committee (FUGS) – Member
Fall 2015 to present	Faculty Operations Committee (FOPS) – Member
Winter 2016 to present	Engineering Faculty Committee – Member
Winter 2009 to Fall 2010	Computing Committee – Member

Professional Activities

Society Memberships and Positions Held

- Member, Association for Computing Machinery (ACM) and ACM Special Interest Group on Information Retrieval (SIGIR).
- SIGIR IRList Editor (mail list moderator), August 2009 to August 2012.

Area Chair / Senior Program Committee Member

- Area Chair, ACM Conference on Information and Knowledge Management (CIKM), 2014.
- Area Chair (Evaluation of IR), ACM Special Interest Group on Information Retrieval (SIGIR) Conference on Research and Development in Information Retrieval, 2014, 2015, 2016, 2017.

Editorial Boards

- Journal of Information Retrieval, July 2014 to present.
- Foundations and Trends in Information Retrieval, February 2014 to present.
- Information Processing and Management, April 2014 to August 2015.
- Book Review Editor, Journal of Information Retrieval, November 2011 to June 2014.

Text Retrieval Conference (TREC) Track Co-Organizer

- TREC 2012 and 2013 CrowdSourcing Track.
- TREC 2010 Relevance Feedback Track.

Workshop Co-Organizer

- SIGIR 2013 Workshop On Modeling User Behavior for Information Retrieval Evaluation.
- CIKM 2013 Living Labs for Information Retrieval Evaluation Workshop.
- SIGIR 2010 Workshop on the Simulation of Interaction for Evaluation of Interactive Retrieval Systems.

Organizing Committee Member

- Co-Organizer, Doctoral Consortium, ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR), 2016.
- Co-Organizer of the 7th annual Symposium on Human-Computer Interaction and Information Retrieval (HCIR 2013).
- Co-Chair of SIGIR 2012 Posters and Demos, Fall 2011-August 2012.
- Publications Chair for the 32nd annual SIGIR Conference, 2009.

Program Committee Member

- ACM Special Interest Group on Information Retrieval (SIGIR) Conference on Research and Development in Information Retrieval, (ac=senior program committee, fp=full papers, ws=workshops, sp=short paper, pp=posters, t=tutorials, and d=demos), 2015 sp, 2014 ac/sp/t, 2013 fp/pp, 2012 fp, 2011 fp/pp/ws, 2010 fp/pp/d, 2009 fp/pp/d, 2008 fp/pp/d, 2007 pp.
- International Conference on the Theory of Information Retrieval (ICTIR), 2015, 2016.
- ACM Conference on Information and Knowledge Management (CIKM), 2011-2013.
- European Conference on IR (ECIR), 2008-2014.
- Symposium on Human-Computer Interaction and Information Retrieval (HCIR), 2013.
- Information Interaction in Context (IIIX), 2008, 2010, 2012, 2014.
- World Wide Web Conference (WWW), Behavioral and Personalization Track, 2011, 2013, lead reviewer: 2015
- Asia Information Retrieval Symposium (AIRS), 2009, 2013.

Journal Reviewer

- ACM Transactions on Information Systems (TOIS), 2010, 2011, 2014, 2016.
- Information Processing and Management (IPM), 2010, 2011, 2014, 2015, 2016.
- Journal of Information Retrieval (IRJ), 2009, 2011.
- ACM Transactions on the Web (TWEB), 2013-2015
- Information Processing Letters, 2014.
- IEEE Knowledge and Data Engineering (KDE), 2009.
- IEEE Transactions on Evolutionary Computation, 1996, 2004.

Monograph Reviewer

- *Information Retrieval Evaluation*, Donna Harmon, Morgan & Claypool, 2011.

Reviewer

- International Conference on the Theory of Information Retrieval (ICTIR), Tutorials, 2013, 2016.
- Human-Computer Interaction and Information Retrieval (HCIR), 2011, 2012.
- SIGCHI, 2011, 2015.
- NAACL-HLT, 2009, 2010.
- World Wide Web Conference (WWW), Demos, 2010.

Workshop Program Committees

- ECIR 2015 Workshop on Supporting Complex Search Tasks.
- SIGIR 2014 Workshop on Gathering Efficient Assessments of Relevance (GEAR).
- SIGIR 2011 EntertainMe Workshop.
- SIGIR 2011 CrowdSourcing for Information Retrieval Workshop.
- ECIR 2011 Workshop on Information Retrieval over Query Sessions.
- WSDM 2011 Workshop on Crowdsourcing for Search and Data Mining.
- SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation.

Other

- SIGIR 40th Anniversary Awards Committee, Member, 2017.
- Reviewer of grant proposals for various european granting agencies, 2014, 2015.
- SIGIR 2014 Doctoral Consortium Mentor.
- SIGIR 2012 Doctoral Consortium Lunch non-student participant.
- Supplied dataset for TREC 2011 crowd source track.
- ICTIR 2009 Session Chair.
- SIGIR conference student volunteer, 2006 and 2007.
- Developed and wrote the 2004 TREC HARD track's metadata guidelines.