

Charles L. A. Clarke
Professor
David R. Cheriton School of Computer Science
University of Waterloo

Education and Professional Qualifications

Doctor of Philosophy, October 1996

Computer Science, University of Waterloo
Thesis: *An Algebra for Structured Text Search*
Supervisor: G. V. Cormack

Master of Mathematics, October 1990

Computer Science, University of Waterloo
Thesis: *Language and Compiler Support for Synchronous Message Passing Architectures*
Supervisor: G. V. Cormack

Bachelor of Science (Honours), May 1986

Mathematics/Computer Science, Memorial University of Newfoundland.
Awarded University Medal for Academic Excellence in Mathematics and Statistics

Professional Engineer, licenced in the Province of Ontario, October 2006

ACM Distinguished Scientist, 2015

Academic and Professional Appointments

2009-present	Professor
2002-2009	Associate Professor
1999-2002	Assistant Professor
	David R. Cheriton School of Computer Science, University of Waterloo
	Visiting Researcher , sabbatical appointment, Facebook, 2016-17
	Director of Software Engineering , University of Waterloo, 2007-2012
	Visiting Researcher , sabbatical appointment, Microsoft, 2006
1996-1999	Assistant Professor
	Department of Electrical and Computer Engineering, University of Toronto
1993-1996	Research Associate (full-time staff position)
	Department of Computer Science, University of Waterloo
1992-1993	Technical Architect
	SHL Systemhouse, Toronto/Los Angeles
1985-1988	Systems Programmer
	Newfoundland Oceans Research and Development Corporation St. John's, Newfoundland

RESEARCH AND SCHOLARSHIP

Areas of Interest

information retrieval, search, data mining, question answering, software tools

Publications

Books and Edited Conference Proceedings

- B1. Shlomo Geva, Andrew Trotman, Peter Bruza, Charles L. A. Clarke, and Kalervo Järvelin (editors). *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR), Gold Coast, Australia, July 2014.
- B2. Stefan Büttcher, Charles L. A. Clarke, and Gordon V. Cormack. *Information Retrieval: Implementing and Evaluating Search Engines*, MIT Press, 606 pages, 2010.
- B3. Charles L. A. Clarke, Norbert Fuhr, Noriko Kando, Wessel Kraaij and Arjen P. de Vries (editors). *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Amsterdam, July 2007.
- B4. Jamie Callan, Charles Clarke, Gordon Cormack, David Hawking and Alan Smeaton (editors). *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Toronto, July 2003.

Journal Articles and Refereed Book Chapters

- J1. Roberto Konow, Gonzalo Navarro, Charles L. A. Clarke, and Alejandro López-Ortiz. Inverted Treaps. To appear in *ACM Transactions on Information Systems*, 2017.
- J2. Charles L. A. Clarke, J. Shane Culpepper, and Alistair Moffat. Assessing Efficiency-Effectiveness Trade-offs in Multi-Stage Retrieval Systems Without Using Relevance Judgments. *Information Retrieval Journal* **19**(4):351-377, August 2016.
- J3. Luchen Tan and Charles L. A. Clarke. A Family of Rank Similarity Measures Based on Maximized Effectiveness Difference. *IEEE Transactions on Knowledge and Data Engineering*, **27**(11):2865-2877, November 2015.
- J4. Azin Ashkan and Charles L.A. Clarke. Location and Query Aware Modeling of Browsing and Click Behavior in Sponsored Search. *ACM Transactions on Intelligent Systems and Technology*, **5**(4):32 pages, 2014.
- J5. Peter B. Golbus, Javed A. Aslam, and Charles L. A. Clarke. Increasing Evaluation Sensitivity to Diversity. *Information Retrieval*, **16**(4):530-555, 2013
- J6. John A. Akinyemi and Charles L.A. Clarke. Fast and Effective Soft Links. *Software Practice and Experience*, **43**(5):577-593, 2013.
- J7. Azin Ashkan and Charles L A. Clarke. Impact of Query Intent and Search Context on Clickthrough Behavior in Sponsored Search. *Knowledge and Information Systems*, **34**(2):425-452, 2013.
- J8. John S. Whissell and Charles L. A. Clarke. Improving Document Clustering using Okapi BM25 Feature Weighting. *Information Retrieval*, **14**(5):466-487, 2011.
- J9. Gordon V. Cormack, Mark D. Smucker, Charles L. A. Clarke. Efficient and Effective Spam Filtering and Re-Ranking for Large Web Datasets. *Information Retrieval*, **14**(5):441-465, 2011.

- J10. Charles L. A. Clarke, Gordon V. Cormack, Thomas R. Lynam, Chris Buckley and Donna Harman. Swapping Documents and Terms. *Information Retrieval*, **12**(6):680-694, December 2009. Condensed version of paper C38 for special issue.
- J11. Stefan Büttcher and Charles L. A. Clarke. Hybrid Index Maintenance for Contiguous Inverted Lists. *Information Retrieval*, **11**(3):175-207, June 2008.
- J12. Charles L. A. Clarke, Gordon V. Cormack, Thomas R. Lynam and Egidio L. Terra. Question Answering by Passage Selection. In *Advances in Open Domain Question Answering*, Tomek Strzalkowski (editor), Springer, 2006.
- J13. Charles L. A. Clarke and Robert C. Good. Schema-Independent Retrieval from Heterogeneous Structured Text. In *Mind Technologies: Humanities Computing and the Canadian Academic Community*, University of Calgary Press, 2006.
- J14. Stefan Büttcher and Charles L. A. Clarke. Adding Full-Text Filesystem Search to Linux. *login: The USENIX Magazine*, 31(3), June 2006.
- J15. G. V. Cormack, C. L. A. Clarke and C. R. Palmer. MultiText Experiments for TREC. In *TREC: Experiments and Evaluation in Information Retrieval*, Donna Harman and Ellen Voorhees (editors), MIT Press, 2005.
- J16. Anthony Cox and Charles Clarke. Multi-Layered Data Modelling. *Electronic Notes in Theoretical Computer Science*, 94:71-79, Elsevier Science, 2004. A preliminary version of this paper was presented at the *First International Workshop on Meta-Models and Schemas for Reverse Engineering*, Victoria, B.C., November, 2003.
- J17. Charles L. A. Clarke and Gordon V. Cormack. Shortest Substring Retrieval and Ranking. *ACM Transactions on Information Systems*, 18(1):44-78, January 2000.
- J18. Charles L. A. Clarke, Gordon V. Cormack and Elizabeth A. Tudhope. Relevance Ranking for One to Three Term Queries. *Information Processing and Management*, 36(2):291-311, 2000. Invited submission. A preliminary version of this paper appears in *Proceeding of the Fifth RIAO Conference*, Montreal, June, 1997.
- J19. Gordon V. Cormack, Charles L. A. Clarke, Christopher R. Palmer and Samuel S. L. To. Passage-Based Query Refinement. *Information Processing and Management*, 36(1):133-153, 2000. Invited submission. A preliminary version of this paper appears in *Proceedings of the Sixth Text REtrieval Conference (TREC-6)*, Gaithersburg, Maryland, November 1997.
- J20. Charles L. A. Clarke and Gordon V. Cormack. On the use of Regular Expressions for Searching Structured Text. *ACM Transactions on Programming Languages and Systems*, 19(3):413-426, 1997.
- J21. Charles L. A. Clarke and David V. Mason. Compacting Garbage Collection can be Fast and Simple. *Software — Practice and Experience*, 26(2):177-194, 1996.
- J22. Charles L. A. Clarke, G. V. Cormack and F. J. Burkowski. An Algebra for Structured Text Search and a Framework for its Implementation. *The Computer Journal*, 38(1):43-56, 1995.

Refereed Full Conference Papers

- C1. Charles L. A. Clarke, Gordon V. Cormack, Jimmy Lin, and Adam Roegiest. Ten Blue Links on Mars. To appear at the *26th World Wide Web Conference*, Perth, Australia, April 2017.
- C2. J. Shane Culpepper, Charles L. A. Clarke and Jimmy Lin. Dynamic Cutoff Prediction in Multi-Stage Retrieval Systems. In *21st Australasian Document Computing Symposium*, pages 17-24, Melbourne, Australia, December 2016. **Best paper award.**

- C3. Carolyn Lamb, Daniel G. Brown, and Charles L.A. Clarke. Evaluating Digital Poetry: Insights from the CAT. In *7th International Conference on Computational Creativity*, pages 60-67, Paris, June 2016
- C4. Carolyn Lamb, Daniel G. Brown and Charlie L.A. Clarke A Taxonomy of Generative Poetry Techniques. In *Bridges Conference*, pages 195-202, Jyväskylä, Finland, August 2016.
- C5. Adam Roegiest, Gordon V. Cormack, Charles L. A. Clarke and Maura R. Grossman. Impact of Surrogate Assessments on High-Recall Retrieval. In *38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 555-564, Santiago, Chile, August 2015.
- C6. Gaurav Baruah, Mark D. Smucker and Charles L. A. Clarke. Evaluating Streams of Evolving News Events. In *38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 675-684, Santiago, Chile, August 2015.
- C7. Carolyn Lamb, Daniel G. Brown and Charles Clarke Can Human Assistance Improve a Computational Poet? In *Bridges Conference*, pages 37-44, Baltimore, July 2015.
- C8. Carolyn Lamb, Daniel G. Brown and Charles L. A. Clarke. Human Competence in Creativity Evaluation. In *6th International Conference on Computational Creativity*, pages 102-109, Park City, Utah, June 2015.
- C9. Charles L. A. Clarke and Mark D. Smucker. Time Well Spent. In *5th Information Interaction in Context Symposium (IIIIX)*, pages 205-214, Regensburg, Germany, August 2014.
- C10. John S. Whissell and Charles L. A. Clarke. Effective Measures for Inter-document Similarity. In *22nd ACM Conference on Information and Knowledge Management*, pages 1361-1370, San Francisco, October 2013.
- C11. Roberto Konow, Gonzalo Navarro, Charles L. A. Clarke, and Alejandro Lòpez-Ortiz. Faster and Smaller Inverted Indices with Treaps. In *36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 193-202, Dublin, July 2013.
- C12. Mark Smucker and Charles L. A. Clarke. Modeling User Variance in Time-Biased Gain. In *6th Symposium on Human-Computer Interaction and Information Retrieval*, Cambridge, Massachusetts, October 2012.
- C13. Mark D. Smucker and Charles L. A. Clarke. Time-Based Calibration of Effectiveness Measures. In *35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 95-104, Portland, August, 2012. **Best paper award.**
- C14. John S. Whissell and Charles L. A. Clarke. Clustering for Semi-Supervised Spam Filtering. In *Proceedings of the 8th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference*, pages 125-134, Perth, Australia, September 2011.
- C15. Charles L. A. Clarke, Nick Craswell, Ian Soboroff, and Azin Ashkan. A Comparative Analysis of Cascade Measures for Novelty and Diversity. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining* pages 75-84, Hong Kong, February 2011.
- C16. Azin Ashkan and Charles L. A. Clarke. On the Informativeness of Cascade and Intent-Aware Effectiveness Measures. In *Proceedings of the 20th International World Wide Web Conference*, pages 407-416, Hyderabad, India, March 2011.
- C17. Benjamin Lafreniere, Andrea Bunt, John S. Whissell, Charles L. A. Clarke, and Michael Terry. Characterizing Large-Scale Use of a Direct Manipulation Application in the Wild, In *Proceedings of the 36th Graphics Interface Conference*, pages 11-18, May 2010.

- C18. Azin Ashkan and Charles L. A. Clarke. Characterizing Commercial Intent. In *Proceedings of the Eighteenth ACM Conference on Information and Knowledge Management*, pages 67-76, Hong Kong, November 2009.
- C19. John Whissell, Charles L. A. Clarke, and Azin Ashkan. Clustering Web Queries In *Proceedings of the Eighteenth ACM Conference on Information and Knowledge Management*, pages 899-908, Hong Kong, 2009.
- C20. Azin Ashkan, Charles Clarke, Eugene Agichtein, and Qi Guo. Estimating Ad Clickthrough Rate through Query Intent Analysis. In *IEEE/WIC/ACM International Conference on Web Intelligence*, pages 222-229, Milan, September 2009.
- C21. Charles L. A. Clarke, Maheedhar Kolla and Olga Vechtomova. An Effectiveness Measure for Ambiguous and Underspecified Queries. In *2nd International Conference on the Theory of Information Retrieval*, pages 188-199, Cambridge, England, September 2009.
- C22. Charles L. A. Clarke, Maheedhar Kolla, Gordon V. Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher and Ian MacKinnon. Novelty and Diversity in Information Retrieval Evaluation. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 659-666, Singapore, 2008.
- C23. Stefan Büttcher and Charles L. A. Clarke. Index Compression is Good, Especially for Random Access. In *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management*, pages 761-770, Lisbon, 2007.
- C24. Charles L. A. Clarke, Eugene Agichtein, Susan Dumais and Ryen W. White. The Influence of Caption Features on Clickthrough Patterns in Web Search. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 135-142, Amsterdam, 2007.
- C25. Stefan Büttcher, Charles L. A. Clarke, Peter C. K. Yeung and Ian Soboroff. Reliable Information Retrieval Evaluation with Incomplete and Biased Judgements. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 63-70, Amsterdam, 2007.
- C26. Stefan Büttcher and Charles L. A. Clarke. A Document-Centric Approach to Static Index Pruning In Text Retrieval Systems. In *15th ACM Conference on Information and Knowledge Management (CIKM)*, pages 182-189, Arlington, Virginia, November 2006.
- C27. Luanne Freund, Charles Clarke and Elaine Toms. Towards Genre classification for IR in the Workplace. In *First Symposium on Information Interaction in Context (IiX)*, Copenhagen, October 2006.
- C28. Stefan Büttcher, Charles L. A. Clarke and Brad Lushman. Hybrid Index Maintenance for Growing Text Collections. In *29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 175-207, Seattle, August 2006.
- C29. Celine Latulipe, Ian Bell, Charles L.A. Clarke and Craig S. Kaplan. symTone: Two-Handed Manipulation of Tone Reproduction Curves. In *Graphics Interface 2006 (GI2006)*, Québec City, June 2006. **Best student paper award.**
- C30. Stefan Büttcher and Charles L. A. Clarke. A Hybrid Approach to Index Maintenance in Dynamic Text Retrieval Systems. In *28th European Conference on Information Retrieval (ECIR 2006)*, London, UK, April 2006. **Best student paper award.**
- C31. Celine Latulipe, Stephen Mann, Charles L.A. Clarke and Craig S.Kaplan. symSpline: Bimanual Symmetric Spline Manipulation. In *2006 Conference on Human Factors in Computing Systems (CHI 2006)*, Montréal, April 2006.

- C32. Stefan Büttcher and Charles L. A. Clarke. A Security Model for Full-Text Search in Multi-User Environments. In *4th USENIX Conference on File and Storage Technologies (USENIX FAST)*, San Francisco, December 2005.
- C33. Celine Latulipe, Craig S. Kaplan and Charles L.A. Clarke. Bimanual and Unimanual Image Alignment: An Evaluation of Mouse-Based Techniques. In *18th Annual ACM Symposium on User Interface Software and Technology (UIST)*, Seattle, October 2005.
- C34. Celine Latulipe, Craig S. Kaplan and Charles L.A. Clarke. Simultaneous Rotation and Translation. In *19th British HCI Group Annual Conference*, British Computer Society, Edinburgh, September 2005.
- C35. Charles L. A. Clarke. Controlling Overlap in Content-Oriented XML Retrieval. In *28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Salvador, Brazil, August 2005.
- C36. Luanne S. Freund, Elaine G. Toms and Charles L.A. Clarke. Modeling Task-Genre Relationships for IR in the Workplace. In *28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Salvador, Brazil, August 2005.
- C37. Charles L. A. Clarke and Egidio L. Terra. Approximating the Top- m Passages in a Parallel Question Answering System. In *13th ACM Conference on Information and Knowledge Management (CIKM)*, Washington DC, November 2004.
- C38. Thomas R. Lynam, Chris Buckley, Charles L. A. Clarke and Gordon V. Cormack. A Multi-System Analysis of Document and Term Selection for Blind Feedback. In *13th ACM Conference on Information and Knowledge Management (CIKM)*, Washington DC, November 2004.
- C39. Egidio Terra and Charles L. A. Clarke. Scoring Missing Terms in Information Retrieval Tasks. In *13th ACM Conference on Information and Knowledge Management (CIKM)*, Washington DC, November 2004.
- C40. Egidio Terra and Charles L.A. Clarke. Fast Computation of Lexical Affinity Models. In *20th International Conference on Computational Linguistics*, pages 50-58, Geneva, Switzerland, July 2004.
- C41. Christopher A. Mennie and Charles L. A. Clarke. Giving Meaning to Macros. In *12th International Workshop on Program Comprehension (IWPC)*, pages 79-88, Bari, Italy, June 2004.
- C42. Charles L. A. Clarke, Philip L. Tilker, Allen Quoc-Luan Tran, Kevin Harris, and Antonio S. Cheng. A Reliable Storage Management Layer for Distributed Information Retrieval Systems. In *12th ACM Conference on Information and Knowledge Management (CIKM)*, New Orleans, November 2003.
- C43. E. L. Terra and Charles Clarke. Frequency Estimates for Statistical Word Similarity Measures. In *2003 Human Language Technology Conference (HLTC)*, Edmonton, Canada, May 2003.
- C44. Anthony Cox and Charles Clarke. Syntactic Approximation Using Iterative Lexical Analysis. In *11th International Workshop on Program Comprehension (IWPC)*, Portland, Oregon, May 2003.
- C45. Chiasen Chung and Charles L. A. Clarke. Topic-Oriented Collaborative Crawling. In *11th ACM Conference on Information and Knowledge Management (CIKM)*, McLean, Virginia, November 2002.
- C46. Anthony Cox and Charles Clarke. Representing and Accessing Extracted Information. In *IEEE International Conference on Software Maintenance (ICSM)*, Florence, Italy, November, 2001.
- C47. Charles L. A. Clarke, Gordon V. Cormack and Thomas R. Lynam Exploiting Redundancy in Question Answering. In *24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New Orleans, September, 2001.

- C48. Susan E. Sim, Charles L. A. Clarke, Ric C. Holt, and Anthony M. Cox. Browsing and Searching Software Architectures. In *IEEE International Conference on Software Maintenance (ICSM)*, Oxford, England, September 1999.
- C49. Gordon V. Cormack, Christopher R. Palmer and Charles L. A. Clarke. Efficient Construction of Large Test Collections. In *21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Melbourne, August, 1998.
- C50. Susan E. Sim, Charles L. A. Clarke and Ric C. Holt. Archetypal Source Code Searching. In *6th IEEE International Workshop on Program Comprehension (IWPC)*, Ischia, Italy, June 1998.
- C51. Charles L. A. Clarke, Gordon V. Cormack and Forbes J. Burkowski. Schema-Independent Retrieval from Heterogeneous Structured Text. In *4th Annual Symposium on Document Analysis and Information Retrieval (SDAIR)*, Las Vegas, Nevada, April 1995.
- C52. F. J. Burkowski, C. L. A. Clarke, S. C. Cowan and G. J. Vreugdenhil. Performance Evaluation of the Sylvan Multiprocessor Architecture. In *Symposium on Experiences with Distributed and Multiprocessor Systems*, Atlanta, March 1991.

Refereed Short Papers and Posters

- S1. Mark D. Smucker and Charles L. A. Clarke. Modeling Optimal Switching Behavior. Short paper in *1st ACM SIGIR Conference on Human Information Interaction and Retrieval*, pages 317-320, Chapel Hill, North Carolina, March 2016.
- S2. Charles L. A. Clarke, Gordon V. Cormack, Jimmy J. Lin, and Adam Roegiest. Total Recall: Blue Sky on Mars. Short paper in *2nd ACM SIGIR International Conference on the Theory of Information Retrieval*, pages 45-48, Newark, Delaware, September 2016.
- S3. Luchen Tan, Adam Roegiest, Jimmy Lin, and Charles L. A. Clarke. An Exploration of Evaluation Metrics for Mobile Push Notifications. Short paper in *39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 741-744, Pisa, Italy, July 2016.
- S4. Luchen Tan, Adam Roegiest, Charles L. A. Clarke and Jimmy Lin. Simple Dynamic Emission Strategies for Microblog Filtering. Short paper in *39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1009-1012, Pisa, Italy, July 2016.
- S5. Adam Roegiest, Luchen Tan, Jimmy J. Lin, and Charles L. A. Clarke. A Platform for Streaming Push Notifications to Mobile Assessors. Short paper in *39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1077-1080, Pisa, Italy, July 2016.
- S6. Luchen Tan, Haotian Zhang, Charles L. A. Clarke and Mark D. Smucker. Lexical Comparison Between Wikipedia and Twitter Corpora by Using Word Embeddings. Short paper in *53rd Annual Meeting of the Association for Computational Linguistics*, pages 657-661, Beijing, July, 2015.
- S7. Seyyed Hadi Hashemi, Charles L. A. Clarke, Adriel Dean-Hall, Jaap Kamps and Julia Kiseleva. On the Reusability of Open Test Collections. Short paper in *38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 827-830, Santiago, Chile, August 2015.
- S8. Adriel Dean-Hall and Charles L.A. Clarke. The Power of Contextual Suggestion. Short paper in *37th European Conference on Information Retrieval Research (ECIR)*, pages 352-357, Vienna, April 2015. To appear.
- S9. Luchen Tan and Charles L.A. Clarke. Succinct Queries for Linking and Tracking News in Social Media. Short paper in *23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM)*, pages 1883-1886, Shanghai, November 2014.

- S10. Adriel Dean-Hall and Charles L. A. Clarke. Assessing Contextual Suggestion. Short paper in *6th International Workshop on Evaluating Information Access (EVIA)*, Tokyo, December 2014.
- S11. John Whissell, and Charles Clarke Classification-Based Clustering Evaluation. Short paper in *IEEE International Conference on Data Mining*, Dallas, December 2013.
- S12. Tetsuya Sakai, Zhicheng Dou and Charles L. A. Clarke. The impact of intent selection on diversified search evaluation. Poster paper in *36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 921-924, Dublin, July 2013.
- S13. Mark D. Smucker and Charles L. A. Clarke. Stochastic Simulation of Time-Biased Gain. Short paper in *21st ACM International Conference on Information and Knowledge Management*, pages 2040-2044, October, 2012.
- S14. Azin Ashkan and Charles L. A. Clarke. Modeling Browsing Behavior for Click Analysis in Sponsored Search. Short paper in *21st ACM International Conference on Information and Knowledge Management*, pages 2015-2019, October, 2012.
- S15. Gobaan Raveendran and Charles L. A. Clarke. Lightweight Contrastive Summarization for News Comment Mining. Poster paper in *35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1103-1104, Portland, August, 2012.
- S16. Kelly Y. Itakura, Charles L. A. Clarke, Shlomo Geva, Andrew Trotman, and Wei Chi Huang. Topical and Structural Linkage in Wikipedia. Short paper in *33rd European Conference on Information Retrieval*, pages 460-465, 2011.
- S17. John A. Akinyemi and Charles L. A. Clarke. Do Subtopic Judgments Reflect Diversity? Poster paper presented at *3rd International Conference on the Theory of Information Retrieval*, pages 309-312, 2011.
- S18. John Akinlabi Akinyemi, Charles L. A. Clarke and Maheedhar Kolla. Towards a Collection-Based Results Diversification. Short paper presented at *9th RIAO Conference*, pages 202-205, Paris, April 2010.
- S19. Kelly Y. Itakura and Charles L. A. Clarke. A Framework for BM25F-based XML Retrieval. Poster paper in *33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 843-844, Geneva, July 2010.
- S20. Qi Guo, Eugene Agichtein, Charles Clarke, and Azin Ashkan. In the Mood to Click? Inferring Searcher Receptiveness to Advertising via Modeling Contextualized Interactions. Short paper in *IEEE/WIC/ACM International Conference on Web Intelligence*, pages 319-324, Milan, September 2009.
- S21. Azin Ashkan and Charles Clarke. Term-Based Commercial Intent Analysis. Poster paper in *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 800-801, Boston, July, 2009.
- S22. Kelly Itakura and Charles Clarke. Using Dynamic Markov Compression to Detect Vandalism in the Wikipedia. Poster paper in *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 822-823, Boston, July, 2009.
- S23. Gordon Cormack, Stefan Büttcher and Charles Clarke. Reciprocal Rank Fusion Outperforms Condorcet and Individual Rank Learning Methods. Poster paper in *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 758-759, Boston, July, 2009.
- S24. Azin Ashkan, Charles L. A. Clarke, Eugene Agichtein, and Qi Guo. Classifying and Characterizing Query Intent. Short paper in *31st European Conference on Information Retrieval*, pages 578-586, Toulouse, France, April, 2009.

- S25. Ryen W. White, Charles L. A. Clarke and Silviu Cucerzan. Comparing Query Logs and Pseudo-Relevance Feedback for Web-Search Query Refinement, Poster paper in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 831-832, Amsterdam, 2007.
- S26. Peter C. K. Yeung, Charles L. A. Clarke and Stefan Büttcher. Improving Retrieval Accuracy by Weighting Document Types with Clickthrough Data. Poster paper in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 759-760, Amsterdam, 2007.
- S27. Peter C.K. Yeung and Luanne Freund and Charles L.A. Clarke. X-Site: A Workplace Search Tool for Software Engineers. Demonstration abstract in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, page 900, Amsterdam, 2007.
- S28. Peter C. K. Yeung, Stefan Büttcher, Charles L. A. Clarke and Maheedhar Kolla. A Bayesian approach for Learning Document Type Relevance. In *29th European Conference on Information Retrieval*, pages 753-756, Rome, April 2007.
- S29. Stefan Büttcher, Charles L. A. Clarke and Brad Lushman. Term Proximity Scoring for Ad-Hoc Retrieval on Very Large Text Collections. Poster paper in *29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Seattle, August 2006.
- S30. Stefan Büttcher and Charles L. A. Clarke. Indexing Time vs. Query Time Trade-offs in Dynamic Information Retrieval Systems. Poster paper in *14th ACM Conference on Information and Knowledge Management (CIKM)*, Bremen, Germany, October 2005.
- S31. Egidio Terra, Kevyn Collins-Thompson, Jamie Callan and Charles L. A. Clarke. The Effect of Document Retrieval Quality on Factoid Question Answering Performance. Poster paper in *27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Sheffield, UK, July 2004.
- S32. Charles L. A. Clarke and Egidio L. Terra. Passage Retrieval vs. Document Retrieval for Factoid Question Answering. Poster paper in *26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Toronto, Canada, July 2003.
- S33. C. L. A. Clarke, G. V. Cormack, M. Laszlo, T. R. Lynam and E. L. Terra. The Impact of Corpus Size on Question Answering Performance. Poster paper in *25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Tampere, Finland, August 2002.
- S34. T. R. Lynam and C. L. A. Clarke and G. V. Cormack. Information Extraction with Term Frequencies. Short paper in *2001 Human Language Technology Conference (HLTC)*, San Diego, 2001.
- S35. Charles Clarke, Anthony Cox and Susan Sim. Searching Program Source with a Structured Text Retrieval System. Poster paper in *22st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Berkeley, August, 1999.
- S36. Gordon V. Cormack, Charles L. A. Clarke, Christopher R. Palmer, and Robert C. Good. The MultiText Retrieval System. Demonstration abstract in *22st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Berkeley, August, 1999.

Selected Workshop Papers

- W1. Seyyed Hadi Hashemi, Jaap Kamps, Julia Kiseleva, Charles L.A. Clarke, and Ellen Voorhees. Overview of the TREC 2016 Contextual Suggestion Track. Presented at the *25th Text REtrieval Conference (TREC)*, Gaithersburg, Maryland, November 2016.

- W2. Adam Roegiest, Gordon V. Cormack, Maura R. Grossman and Charles L.A. Clarke. TREC 2016 Total Recall Track Overview. Presented at the *25th Text REtrieval Conference (TREC)*, Gaithersburg, Maryland, November 2016.
- W3. Adriel Dean-Hall, Charles L. A. Clarke, Jaap Kamps and Julia Kiseleva. Online Evaluation of Point-Of-Interest Recommendation Systems. In *First International Workshop on Supporting Complex Search Tasks*, ECIR workshop, Vienna, Austria, March 29, 2015.
- W4. Adriel Dean-Hall, Charles L. A. Clarke, Jaap Kamps, Julia Kiseleva, and Ellen Voorhees. Overview of the TREC 2015 Contextual Suggestion Track. Presented at the *24th Text REtrieval Conference (TREC)*, Gaithersburg, Maryland, November 2015.
- W5. Adam Roegiest, Gordon V. Cormack, Charles L.A. Clarke, and Maura R. Grossman. TREC 2015 Total Recall Track Overview. Presented at the *24th Text REtrieval Conference (TREC)*, Gaithersburg, Maryland, November 2015.
- W6. Adriel Dean-Hall, Charles L. A. Clarke, Jaap Kamps, Paul Thomas, and Ellen Voorhees. Overview of the TREC 2014 Contextual Suggestion Track. Presented at the *23rd Text REtrieval Conference (TREC)*, Gaithersburg, Maryland, November 2014.
- W7. Adriel Dean-Hall, Charles L. A. Clarke, Jaap Kamps, Paul Thomas, Nicole Simone, and Ellen Voorhees. Overview of the TREC 2013 Contextual Suggestion Track. Presented at the *22nd Text REtrieval Conference*, Gaithersburg, Maryland, November 2013.
- W8. Kevyn Collins-Thompson, Paul Bennett, Fernando Diaz, Charles L. A. Clarke, and Ellen M. Voorhees. TREC 2013 Web Track Overview. Presented at the *22nd Text REtrieval Conference*, Gaithersburg, Maryland, November 2013.
- W9. Charles L. A. Clarke. Nugget-based computation of graded relevance. Presented at the *5th International Workshop on Evaluating Information Access (E VIA)*, Tokyo, June 2013.
- W10. Adriel Dean-Hall and Charles L. A. Clarke. Evaluating contextual suggestion. Presented at the *5th International Workshop on Evaluating Information Access (E VIA)*, Tokyo, June 2013.
- W11. Adriel Dean-Hall, Charles L. A. Clarke, Jaap Kamps, Paul Thomas and Ellen M. Voorhees. Overview of the TREC 2012 Contextual Suggestion Track. Presented at *21st Text REtrieval Conference*, November 2012.
- W12. Charles L. A. Clarke, Nick Craswell and Ellen M. Voorhees. Overview of the TREC 2012 Web Track. Presented at *21st Text REtrieval Conference*, November 2012.
- W13. Mark D. Smucker and Charles L. A. Clarke. The Fault, Dear Researchers, is not in Cranfield, But in our Metrics, that they are Unrealistic. Position paper in *2nd European Workshop on Human-Computer Interaction and Information Retrieval*, pages 11-12, Nijmegen, August, 2012.
- W14. Charles L. A. Clarke, Nick Craswell, Ian Soboroff, and Ellen M Voorhees. Overview of the TREC 2011 Web Track. Presented at *20th Text REtrieval Conference*, November 2011.
- W15. Azin Ashkan, John S. Whissell and Charles L.A. Clarke. Unsupervised Learning of Result Page Context for Clickthrough Analysis in Sponsored Search. Presented at 5th Annual International Workshop on Data Mining and Audience Intelligence for Online Advertising, San Diego, August 2011.
- W16. Charles L. A. Clarke, Nick Craswell, Ian Soboroff, and Gordon V. Cormack. Overview of the TREC 2010 Web Track. Presented at *19th Text REtrieval Conference*, November 2010.
- W17. Kelly Y. Itakura and Charles L. A. Clarke. University of Waterloo at INEX 2009: Ad Hoc, Book, Entity Ranking, and Link-the-Wiki Tracks. Presented at *8th Workshop of the Initiative for the Evaluation of XML Retrieval*, December 2009.

- W18. Charles L. A. Clarke, Nick Craswell, and Ian Soboroff. Overview of the TREC 2009 Web Track. Presented at *18th Text REtrieval Conference*, November 2009.
- W19. Kelly Y. Itakura and Charles L.A. Clarke. University of Waterloo at INEX2008: Adhoc, Book, and Link-the-Wiki Tracks. In *7th Workshop of the Initiative for the Evaluation of XML Retrieval*, December 2008.
- W20. Azin Ashkan, Charles L. A. Clarke, Eugene Agichtein, and Qi Guo. Characterizing Query Intent From Ad Clickthrough Data. In *1st Workshop on Information Retrieval in Advertising*, SIGIR Workshop, Singapore, July 2008.
- W21. Maheedhar Kolla, Olga Vechtomova, and Charles L. A. Clarke. Nugget-based Framework to Support Graded Relevance Assessments. In *Beyond Binary Relevance: Preferences, Diversity, and Set-Level Judgments*, SIGIR Workshop, Singapore, July 2008
- W22. S. Büttcher, C. L. A. Clarke, G. V. Cormack and T. R. Lynam. MultiText Legal Experiments at TREC 2007. Presented at *16th Text REtrieval Conference*, November 2007.
- W23. Kelly Itakura and Charles L. A. Clarke. From Passages into Elements in XML Retrieval. In *SIGIR 2007 Workshop on Focused Retrieval*, Amsterdam, 2007.
- W24. Stefan Büttcher, Charles L. A. Clarke and Peter C. K. Yeung. Index Pruning and Result Reranking: Effects on Adhoc Retrieval and Named Page Finding. Presented at the *15th Text REtrieval Conference (TREC 2006)*. Gaithersburg, Maryland, November 2006.
- W25. Stefan Büttcher, Charles L. A. Clarke and Ian Soboroff. The TREC 2006 Terabyte Track. Presented at the *15th Text REtrieval Conference (TREC 2006)*. Gaithersburg, Maryland, November 2006.
- W26. Stefan Büttcher and Charles L. A. Clarke. Efficiency vs. Effectiveness in Terabyte-Scale Information Retrieval. Presented at the *14th Text REtrieval Conference (TREC 2005)*. Gaithersburg, Maryland, November 2005.
- W27. Charles L. A. Clarke, Falk Scholer and Ian Soboroff. The TREC 2005 Terabyte Track. Presented at the *14th Text REtrieval Conference (TREC 2005)*. Gaithersburg, Maryland, November 2005.
- W28. Egidio Terra and Charles L. A. Clarke. Comparing Query Formulation and Lexical Affinity Replacements in Passage Retrieval. In *ELECTRA: Methodologies and Evaluation of Lexical Cohesion Techniques in Real-World Applications*, SIGIR Workshop, Salvador, Brazil, August 2005.
- W29. Charles Clarke, Nick Craswell and Ian Soboroff. Overview of the TREC 2004 Terabyte Track, Presented at the *2004 Text REtrieval Conference (TREC 2004)*, Gaithersburg, Maryland, November 2004.
- W30. Charles L. A. Clarke and Philip L. Tilker. MultiText Experiments for INEX 2004. In *Advances in XML Information Retrieval*, 3rd Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2004, Schloss Dagstuhl, Germany, December 2004. Published as Lecture Notes in Computer Science, Volume 3493, Springer, 2005.
- W31. Stefan Büttcher, Charles L. A. Clarke and Gordon V. Cormack. Domain-Specific Synonym Expansion and Validation for Biomedical Information Retrieval. Presented at the *2004 Text REtrieval Conference (TREC 2004)*, Gaithersburg, Maryland, November 2004.
- W32. David L. Yeung, Charles L. A. Clarke, Gordon V. Cormack, Thomas R. Lynam and Egidio L. Terra. Task-Specific Query Expansion. Presented at the *2003 Text REtrieval Conference (TREC 2003)*, Gaithersburg, Maryland, November 2003.
- W33. C. L. A. Clarke and G. V. Cormack and G. Kemkes and M. Laszlo and T. R. Lynam and E. L. Terra and P. L. Tilker. Statistical Selection of Exact Answers. Presented at the *2002 Text REtrieval Conference (TREC 2002)*, Gaithersburg, Maryland, November 2002.

- W34. Anthony Cox and Charles Clarke. Relocating XML Elements from Preprocessed to Unprocessed Code. In *Proceedings of the 10th International Workshop on Program Comprehension (IWPC)*, Paris, June 2002.
- W35. C. L. A. Clarke, G. V. Cormack, T. R. Lynam, C. M. Li and G. L. McLearn. Web Reinforced Question Answering. Presented at the *2001 Text REtrieval Conference (TREC 2001)*, Gaithersburg, Maryland, November 2001.
- W36. Anthony Cox and Charles Clarke. A Comparative Evaluation of Techniques for Syntactic Level Source Code Analysis. In *Proceedings of the Seventh Asia-Pacific Software Engineering Conference*, Singapore, December, 2000.
- W37. C. L. A. Clarke, G. V. Cormack, D. I. E. Kisman and T. R. Lynam. Question Answering by Passage Selection. Presented at the *Ninth Text REtrieval Conference (TREC-9)*, Gaithersburg, Maryland, November 2000.
- W38. Anthony Cox and Charles Clarke. A Functional Approach to Complex Retrieval Tasks. In *9th International Workshop on Functional and Logic Programming*, Benicassim, Spain, September 2000.
- W39. G.V. Cormack, C.L.A. Clarke, C.R. Palmer and D.I.E. Kisman. Fast Automatic Passage Ranking. Presented at *8th Text REtrieval Conference (TREC-8)*, Gaithersburg, Maryland, November 1999.
- W40. Anthony Cox, Charles Clarke and Susan Sim. A Model Independent Source Code Repository. In *CASCON'99*, Toronto, November 1999.
- W41. Gordon V. Cormack, Christopher R. Palmer, Michael Van Biesbrouck and Charles L. A. Clarke. Deriving Very Short Queries for High Precision and Recall. Presented at the *7th Text REtrieval Conference (TREC-7)*, Gaithersburg, Maryland, November 1998.
- W42. Gordon V. Cormack, Charles L. A. Clarke, Christopher R. Palmer and Samuel S. L. To. Passage-Based Refinement. Presented at the *6th Text REtrieval Conference (TREC-6)*, Gaithersburg, Maryland, November 1997.
- W43. Charles L. A. Clarke and Gordon V. Cormack. Interactive Substring Retrieval. Presented at the *5th Text REtrieval Conference (TREC-5)*, Gaithersburg, Maryland, November 1996.
- W44. R. C. Good, G. V. Cormack, C. L. A. Clarke and D. J. Taylor. A Robust Storage Systems Architecture. In *8th International Conference on Computing and Information (ICCI)*, June 1996.
- W45. Charles L. A. Clarke, Gordon V. Cormack and Forbes J. Burkowski. Shortest Substring Ranking. Presented at *4th Text REtrieval Conference (TREC-4)*, Gaithersburg, Maryland, November 1995.

Other Selected Contributions

- X1. Jimmy Lin, Charles L.A. Clarke, and Gaurav Baruah. Searching from Mars. *IEEE Internet Computing* **20**(1):78-82, 2016.
- X2. Grace Hui Yang, Ian Soboroff, Li Xiong, Charles L. A. Clarke, and Simson L. Garfinkel. Privacy-Preserving IR 2016: Differential Privacy, Search, and Social Media. Workshop at *39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1247-1248, Pisa, Italy, July 2016.
- X3. Alessandro Moschitti, Lluís Màrquez, Preslav Nakov, Eugene Agichtein, Charles L. A. Clarke, and Idan Szpektor. Web Question Answering Beyond Factoids. Workshop at *39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1251-1252, Pisa, Italy, July 2016.

- X4. Charles L. A. Clarke, Mark D. Smucker, and Emine Yilmaz. IR Evaluation: Modeling User Behavior for Measuring Effectiveness. Tutorial at *38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1117-1120, Santiago, Chile, August 2015.
- X5. Eugene Agichtein, David Carmel, Charles L. A. Clarke, Praveen Paritosh, Dan Pelleg and Idan Szpektor. Web Question Answering: Beyond Factoids. Workshop at *38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1143, Santiago, Chile, August 2015.
- X6. M.-Dyaa Albakour, Craig Macdonald, Iadh Ounis, Charles L. A. Clarke, and Veli Bicer. Information Access in Smart Cities (i-ASC). Workshop report in *36th European Conference on Information Retrieval Research (ECIR)*, pages 810-814, Amsterdam, April 2014.
- X7. Charles L. A. Clarke, Luanne Freund, Mark D. Smucker, and Emine Yilmaz. SIGIR 2013 workshop on modeling user behavior for information retrieval evaluation. In *36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, page 1134, Dublin, July 2013.
- X8. Andrew Trotman, Charles L. A. Clarke, Iadh Ounis, Shane Culpepper, Marc-Allen Cartright and Shlomo Geva. Open source information retrieval: A report on the SIGIR 2012 workshop. *SIGIR Forum* **46**(2):95-101, 2012.
- X9. Craig Macdonald, Jun Wang and Charles L. A. Clarke. The 2nd International Workshop on Diversity in Document Retrieval. Workshop report in *5th International Conference on Web Search and Web Data Mining*, Seattle, February 2012.
- X10. Nicholas J. Belkin, Charles L. A. Clarke, Ning Gao, Jaap Kamps and Jussi Karlgren. Report on the SIGIR Workshop On “entertain me”: Supporting Complex Search Tasks. *ACM SIGIR Forum*, **45**(2):51-59, December 2011.
- X11. Craig Macdonald, Charles Clarke, and Jun Wang. The 1st International Workshop on Diversity in Document Retrieval, *ACM SIGIR Forum*, **45**(2):87-93, December 2011.
- X12. Charles L. A. Clarke. Web Question Answering. In *Encyclopedia of Database Systems*, Ling Liu and M. Tamer Özsu (editors), Springer, 2008.
- X13. Jamie Callan, James Allan, Charles L. A. Clarke, Susan Dumais, David A. Evans, Mark Sanderson and ChengXiang Zhai. Meeting of the MINDS: An information retrieval research agenda. *ACM SIGIR Forum*, **41**(2):25-34, December 2007.
- X14. Charles L. A. Clarke, Nick Craswell and Ian Soboroff. The TREC Terabyte Retrieval Track *SIGIR Forum*, **39**(1):25, 2005.
- X15. Charles L. A. Clarke, Gordon V. Cormack, Christopher R. Palmer. An Overview of MultiText, *SIGIR Forum*, **32**(2):14-15 (1998). Abstract of a presentation given at the SIGIR '98 workshop on *Information Retrieval: Theory into Practice*, August 1998.

Invited Talks and Presentations (partial list)

March 2016	1st Conference on Human Information Interaction and Retrieval, UNC (invited tutorial)
December 2015	7th Forum for Information Retrieval Evaluation, Gandhinagar, India
August 2015	9th Russian Summer School in Information Retrieval, St. Petersburg
February 2015	WSDM tutorial on Dynamic Information Retrieval Modeling, Shanghai
February 2014	LTI Colloquium, CMU
September 2013	9th European Summer School in Information Retrieval, Granada, Spain
June 2013	NTCIR, National Institute of Informatics, Tokyo (keynote)
February 2013	School of Computing Science, University of Glasgow
December 2012	17th Australasian Document Computing Symposium (keynote)
June 2010	NTCIR, National Institute of Informatics, Tokyo
July 2009	SIGIR Workshop on Redundancy, Diversity and Interdependent Document Relevance (keynote)
December 2008	Microsoft Research, Cambridge
May 2008	Department of Mathematics and Computer Science, Emory University
April 2008	13th Search Engine Meeting, Boston (keynote)
October 2007	UW-IEEE Workshop on Multimedia Mining and Knowledge Discovery
June 2007	ORION Ontario R&D Summit
March 2006	The Information School, University of Washington
May 2005	IBM Toronto Software Lab
April 2005	School of Computer Science, Dalhousie University
February 2005	Microsoft Research, Redmond, Washington
December 2003	4th Dutch-Belgian Information Retrieval Workshop (keynote)
May 2002	Mind Technologies Session, Congress of the Social Sciences and Humanities, Toronto
November 2001	Canadian Digital Libraries Symposium, Toronto
June 2001	IFIP 2.4 working group meeting, San Miniato, Italy
March 2001	CITO Research Review on Software Technology and Distributed Systems
May 2000	IFIP 2.4 working group meeting, Delft, Netherlands

TEACHING ACTIVITIES

Graduate Student Supervision

Current students

Carolyn Lamb (Dan Brown, co-supervisor), Ph.D. expected 2017

Topic: *Found art in social media.*

Amira Ghenai, Ph.D. expected 2020

Topic: *Information retrieval in the medical domain.*

Alexandra Vtyurina, Ph.D. expected 2020

Topic: *Complex question answering.*

Shaikh Quader, M.Math. expected 2017

Topic: *Enterprise search.*

Doctoral students, completed

Luchen Tan, Ph.D. 2016.

Tracking Events in Social Media.

Gaurav Baruah (Mark Smucker, co-supervisor), Ph.D. 2016.

Filtering News from Document Streams.

Azin Ashkan, Ph.D. 2013.

Characterizing User Search Intent and Behaviour for Click Analysis in Sponsored Search.

Maheedhar Kolla (Olga Vechtomova, co-supervisor), Ph.D. 2012.

Novelty and Diversity in Information Retrieval Evaluation.

John Whissell, M.Math. 2006, Ph.D. 2012.

Evaluating Clusterings by Estimating Clarity, Ph.D. thesis

Significant Feature Clustering, M.Math. thesis (Chrysanne DiMarco, co-supervisor)

John Akinyemi, Ph.D. 2012.

Similarity and Diversity in Information Retrieval.

Kelly Itakura, Ph.D. 2010.

Focused Retrieval.

Stefan Büttcher, Ph.D. 2007.

Multi-User File System Search

- Faculty of Mathematics Award for Outstanding Achievement in Graduate Studies

- University Award for Exceptional Teaching by a Student

Celine Latulipe (Craig Kaplan, co-supervisor), Ph.D. 2006.

A Symmetric Interaction Model for Bimanual Input

- Faculty of Mathematics Award for Outstanding Achievement in Graduate Studies

- Alain Fournier Ph.D. Thesis Award

Egidio L. Terra, Ph.D. 2004.

Lexical Affinities and Language Applications.

Tony Cox (G. V. Cormack, co-supervisor), Ph.D. 2002.

A Source-Based Approach to Representing and Managing Information Extracted by Program Analysis.

Master's students, completed

- Adriel Dean-Hall**, M.Math. 2014
Evaluating Contextual Suggestion.
- Younos Abounaga**, M.Math. 2013
Efficient Temporal Synopsis of Social Media Streams.
- Gobaan Raveendran**, M.Math. 2013
Efficiently Crawling, Collecting, and Condensing News Comments.
- Mohammed Alliheedi** (Chrysanne DiMarco, co-supervisor), M.Math. 2012.
Multi-document Summarization Using Rhetorical Information
- Hani Nikkhoo**, M.Math. 2011.
The Impact of Near-Duplicate Documents on Information Retrieval Evaluation
- Peter Yeung**, M.Math. 2007.
Weighting Document Genre in Enterprise Search
- Dimitri Mostinski** (Pascal Poupart, co-supervisor), M.Math. 2007.
Dynamic Factored Particle Filtering for Context-Specific Correlations
- Archana Adma**, M.Math. 2006.
iTalk — Index Your Talk.
- Sandy Graham**, M.Math. 2006 (by essay).
- Philip Tilker**, M.Math. 2005.
A Reliable Storage Management Layer for a Distributed Information Retrieval System.
- Chris Mennie**, M.Math. 2004.
Giving Meaning to Macros.
- John Johansen**, M.Math. 2004.
Delta Pointers.
- Feng Zhao**, M.Math. 2002.
Exploring Altruistic Links in a Web Crawler.
- Chiasen Chung**, M.Math. 2002.
Topic-oriented Collaborative Web Crawling.
- Asim Elahi**, M.Math. 2001 (by essay).
- Alan Tran**, University of Toronto M.Sc. 2000.
A Network Management Facility for a Fault-Tolerant Distributed Information Retrieval System
- Susan Sim** (R. C. Holt, co-supervisor.), University of Toronto M.Sc. 1998.
Supporting Multiple Program Comprehension Strategies During Software Maintenance.
- Tony Cheng**, University of Toronto M.A.Sc. 1998.
A Low-Cost Robust Distributed Storage System for Large Information Retrieval Systems.

Thesis Committee Member (internal)

Aiman L. Al-Harbi (Ph.D. 2016), Andrew Kane (Ph.D. 2014), Rakesh Guttikonda (M.Math. 2014), Kaheer Suleman (M.Math. 2014), Jack Thomas (M.Math. 2014), Francisco Claude (Ph.D. 2013), Le Li (M.Math. 2013), Claus Strommer (Ph.D. 2011), Ting Liu (M.Math. 2011), C P Jethani (M.Math. 2011), Hussein Hirjee (M.Math. 2010), Rehan Rauf (M.Math. 2010), Jakub Gawryjolek (M.Math. 2009), Rana Alshaar (M.A.Sc. 2009), Yuan Lin (Ph.D. 2008), Radoslav Radoulov (M.Math. 2008), Jeff Dicker (M.Math. 2008), Kun Cen (M.A.Sc. 2008), Brad Lushman (M.Math. 2002, Ph.D. 2007), Jingwei Wu (Ph.D. 2006) Ahmed E Hassan (Ph.D. 2005), Jiongxiang Chen (Ph.D. 2005), Allan Menezes (M.Math. 2005), Luke Andrew Tanur (M.Math. 2005), Rick Bodner (Ph.D., Toronto, 2004), Jiongxiang Chen (Ph.D. 2004), Jack S. Chi (M.Math. 2004), Rodolfo Gabriel Esteves Jaramillo (M.Math. 2004), Yuato Guo (M.Math. 2004), Xuhu Li (M.Math. 2004), Fuchun Peng (Ph.D. 2003), Yingwei Wang (Ph.D. 2003), Lijie Zou (M.Math. 2003), Vlado Keselj (Ph.D. 2002), Reem Al-Halimi (Ph.D. 2002), Sunny K. S. Lam (M.Math. 2002), Thomas Lynam (M.Math. 2002), Maryam Aamir Khan (M.Math. 2002), Greg McLearn (M.Math. 2002), Steven Engels (M.Math. 2002), Ian Bull (M.Math. 2002), Zarrin Langari (M.Math. 2001), Rayan Yahfoufi (M.Math. 2001), Stergios Anastasiadis (Ph.D., Toronto, 2001), Michal Ostrowski (M.Math. 2000), Michael Van Biesbrouck (M.Math. 1999), Larry Baptiste (M.A.Sc., Toronto, 1998), Kevin Gior (M.A.Sc., Toronto, 1998), Stephen Green (Ph.D., Toronto, 1998), Bing He (M.Sc., Toronto, 1998), Tony D. Lindsay (M.A.Sc., Toronto, 1998), Tomi Poutanen (M.A.Sc., Toronto, 1998), Nikos Koudas (Ph.D., Toronto, 1997), Dattatraya H. Kulkarni (Ph.D., Toronto, 1997), Sudarson Tandri (Ph.D., Toronto, 1997), Govindan Ravindran (Ph.D., Toronto, 1997), Kelvin Shek Yiu (M.A.Sc., Toronto, 1997), Gene Golovchinsky (Ph.D., Toronto, 1996), Mark Stephen Lawford (Ph.D., Toronto, 1996).

Undergraduate Student Supervision

Nicole Simone, Co-op Student Employee, May.-Aug. 2013
Cat Mercer, Co-op Student Employee, Sept.-Dec. 2012
Michael Chong, Co-op Student Employee, Sept.-Dec. 2011
Ryan King, Co-op Student Employee, Sept.-Dec. 2011 and May-Aug. 2012
William Song, Undergraduate Research Associate, May-Aug. 2011
Trevor Fairey, Co-op Student Employee, May-Aug. 2010
Yubin Kim, Undergraduate Research Associate, May-Aug. 2010
Youngki Chang, Undergraduate Research Associate, May-Aug. 2007
James Rupke, Co-op Student Employee, May-Aug. 2007
Peter Yeung, Co-op Student Employee, May-Aug. 2005

While at the University of Toronto I supervised 26 undergraduate theses and final-year design projects. In 1999, two of my students (**Zhao Ping Tang** and **Chuan Zhang**) received the departmental award for the best final-year design project.

Courses Taught

Term	Title	Level	Class Size
Fall 2015	CS848: Search and Recommendation	G	15
Fall 2014	CS137: Programming Principles	UG	138
Fall 2013	CS137: Programming Principles	UG	152
Spring 2013	CS247: Software Abstraction and Specification	UG	94
Fall 2011	CS848: Information Retrieval	G	9
Fall 2011	CS137: Programming Principles	UG	127
Spring 2011	SE490: Design Project I	UG	75
Fall 2010	CS848: Information Retrieval	G	23
Fall 2010	CS137: Programming Principles	UG	136
Fall 2009	CS137: Programming Principles	UG	116
Fall 2008	CS137: Programming Principles	UG	121
Fall 2007	CS798: Information Retrieval	G	31
Winter 2007	CS452: Real-Time Programming	UG/G	11
Fall 2005	CS133: Developing Programming Principles	UG	114
Fall 2005	CS133: Developing Programming Principles	UG	82
Fall 2004	CS452: Real-Time Programming	UG/G	42
Spring 2004	CS350/353: Operating Systems	UG	74
Spring 2004	CS350/353: Operating Systems	UG	77
Fall 2003	CS452: Real-Time Programming	UG/G	22
Spring 2003	CS848: Automatic Question Answering	G	30
Winter 2003	CS342: Concurrent Programming	UG	101
Winter 2003	CS452: Real-Time Programming	UG/G	7
Spring 2002	CS342: Concurrent Programming	UG	105
Spring 2002	CS342: Concurrent Programming	UG	97
Winter 2002	CS444: Compiler Construction	UG/G	40
Spring 2001	CS452: Real-Time Programming	UG/G	30
Winter 2001	CS444: Compiler Construction	UG/G	34
Winter 2001	CS130: Developing Programming Principles	UG	80
Spring 2000	CS452: Real-Time Programming	UG/G	28
Winter 2000	CS748R: Information Retrieval	G	24
Spring 1999	ECE1765: File Structures and Storage Systems	G	18 (Toronto)
Fall 1998	CSC442: Software Engineering	UG	138 (Toronto)
Fall 1998	APS105: Computer Fundamentals	UG	140 (Toronto)
Spring 1998	ECE1765: File Structures and Storage Systems	G	11 (Toronto)
Spring 1998	ECE544: Large Software Systems	UG	12 (Toronto)
Fall 1997	CSC442: Software Engineering	UG	104 (Toronto)
Spring 1997	ECE544: Large Software Systems	UG	13 (Toronto)
Fall 1996	ECE544: Special Topics in Software Engineering	UG	3 (Toronto)

SERVICE

Internal service

- 2015-2016 Director of Undergraduate Studies, School of Computer Science
- 2013-2014 Associate Director of Undergraduate Studies, School of Computer Science
- 2012-2013 Accreditation Representative for Software Engineering, University of Waterloo
- 2007-2012 Director of the Software Engineering Program, University of Waterloo
- 2004-2005 Graduate Student Advocate, School of Computer Science
- 2004-2005 Undergraduate Academic Plans Committee, School of Computer Science
- 2002-2004 Advisory Committee, Computer Science Computing Facility
- 2001-2004 Computing Advisory Committee, Faculty of Mathematics.
- 1998-1999 Student-Staff Committee, Department of Electrical and Computer Engineering, University of Toronto.
- 1997-1999 Recruiting Committee, Department of Electrical and Computer Engineering, University of Toronto.
- 1997-1999 Coach and faculty advisor for Toronto's ACM International Programming Contest teams. In 1999 one team advanced to the world finals in Eindhoven.

External service

- Co-editor-in-chief, *Information Retrieval Journal*
- General co-chair, *NTCIR Conference Series*, Tokyo, Japan
- Editorial board:
 - ACM Books*
 - Foundations and Trends in Information Retrieval*
 - Springer Informational Retrieval Series*
- Chair, *ACM Special Interest Group on Information Retrieval*, 2013-20116
- Steering committee, *Conference on Information and Knowledge Management*, 2010-2015
- Program co-chair, *ACM SIGIR Conference*, 2014
- Program co-chair, *ACM SIGIR Conference*, 2007
- General co-chair, *ACM SIGIR Conference*, 2003
- Poster chair, *International World Wide Web Conference*, 1999
- Text REtrieval Conference*, Track Coordinator:
 - Total Recall Track Coordinator, 2015-2016.
 - Contextual Suggestion Track Coordinator, 2012-2016.

Web Track Coordinator, 2009-2013.
Terabyte Track Coordinator, 2003-2006.

Workshop organizer:

SIGIR Workshop on Privacy-Preserving IR, 2016
SIGIR Workshop on Web Question Answering, 2015, 2016
ECIR Workshop on Information Access in Smart Cities, 2014
SIGIR Workshop on Modeling User Behavior for IR Evaluation, 2013
SIGIR Workshop on Open Source Information Retrieval, 2012
WSDM Workshop on Diversity in Document Retrieval, 2012
ECIR Workshop on Diversity in Document Retrieval, 2011
SIGIR Workshop on Complex Search Tasks, 2011

Meta-reviewer:

ACM SIGIR Conference, 2002, 2004-2006, 2009-2013, 2015-2016
ACM International Conference on Web Search and Data Mining, 2009, 2012, 2014-2016
International World Wide Web Conference, 2012
HLT-NAACL Human Language Technology Conference, 2003, 2004, 2009

Program committee member:

ACM SIGIR Conference, 2001-2016
ACM CIKM, 2001-2005, 2007-2009, 2013, 2015
ACM KDD, 2015
International Conference on the Theory of Information Retrieval, 2013
European Conference on Information Retrieval, 2007-2010
International World Wide Web Conference, 2008, 2009, 2011, 2012
String Processing and Information Retrieval Symposium, 2008, 2011, 2015
Recherche d'Information Assistée par Ordinateur, 2010
Asia Information Retrieval Symposium, 2008, 2013, 2014
Network and Distributed System Security Symposium, 2007, 2008
Human Language Technology Conference/EMNLP, 2005
Pacific Association for Computational Linguistics Conference, 2005
International Joint Conference on Natural Language Processing, 2005

Reviewer:

ACM Transactions on Information Systems
Information Retrieval
Information Processing and Management
Journal of the American Society for Information Science and Technology
IEEE Transactions on Knowledge and Data Engineering
ACM Computing Surveys
ACM Transactions on the Web
VLBD Journal
Software Practice and Experience
Information Systems
Information Sciences
Computational Intelligence
Knowledge and Information Systems

Grant reviewer:

Swiss National Science Foundation

Netherlands Organisation for Scientific Research
Natural Sciences and Engineering Research Council, Canada
Mitacs, Canada
U.S. National Science Foundation
Hong Kong Research Grants Council
Dutch National Science Foundation
Israel Science Foundation

Ph.D. external examiner:

Julia Kiseleva, University of Eindhoven, 2016
Matt Crane, University of Otago, 2015
Arbi Bouchoucha, University of Montreal, 2015
Peter Golbus, Northeastern, 2014
Falah Al-Akashi, University of Ottawa, 2014
Rodrygo Santos, University of Glasgow, 2013
Mir Sadek Ali, University of Toronto, 2012
William Webber, University of Melbourne, Australia, 2010
Yohannes Tsegay, RMIT University, Melbourne, Australia, 2009
Yang Liu, York University, Toronto, Canada, 2009.
Muath Alzghool, University of Ottawa, 2009
Thanh Tin Tang, Australian National University, 2006.
Dirk Bahle, RMIT University, Melbourne, Australia, 2003.
Christof Monz, University of Amsterdam, 2003.