

ROSIE JONES, PhD

28 Valentine St
Cambridge, MA 02139

Phone: (626)354-5663
Email: rosie.jones@acm.org
Web: <http://www.cs.cmu.edu/~rosie/>

- SUMMARY** Senior scientist and leader with over twenty years of language technologies and machine learning experience. Fourteen years of experience developing practical language technologies in industry. Extensive experience developing and working on research programs in industry and conducting technical transfer. Fourteen awarded patents, over thirty academic publications, and major industry product contributions in sponsored and algorithmic search. Thought leader in industry setting.
- RESEARCH INTERESTS** Natural language processing, text mining, user search behavior, web search, speech applications, conversational agents, computational advertising.
- EDUCATION**
- Carnegie Mellon University**, Pittsburgh, Pennsylvania USA
Ph.D. Language and Information Technologies, 2005
Thesis title: *Learning to Extract Entities from Labeled and Unlabeled Text*
Committee: Tom Mitchell (chair), Alex Hauptmann, Roni Rosenfeld and Ellen Riloff
- Carnegie Mellon University**, Pittsburgh, Pennsylvania USA
M.S. Computational Linguistics, 1997
- University of Sydney**, Sydney, AUSTRALIA
B.Sc (Hons) majoring in Computer Science, 1994
- EXPERIENCE**
- Principal Scientist April, 2013–August 2015
Microsoft Cambridge, MA
Conducted research on conversational understanding systems. Mined big data sets and provided actionable insights to improve algorithms. Conducted experiments. Mentored research engineers. Developed models for production systems. Left for extended maternity leave.
- Director of Data Science January, 2013–April, 2013
MediaMath Cambridge, MA
Managed team of scientists and analysts working on machine learning for computational advertising. MediaMath acquired Akamai’s Advertising Decision Services in January 2013.
- Director of Computational Advertising June, 2010–January, 2013
Akamai Technologies Cambridge, MA
Built and managed team of scientists and analysts working on cutting edge computational advertising algorithms and data analysis. Provided scientific leadership for engineering and product on strategic directions and metrics. Oversaw validation of A/B testing infrastructure and live tests of machine learned models. Worked closely with engineering to put new models into production.
- Senior Research Scientist January, 2004–May 2010
Yahoo! Inc Burbank, CA; Cambridge, MA
Conducted research in information retrieval with applications to web and sponsored search. Used machine learning and statistical analysis of text data to design and run experiments, test hypotheses leading to algorithm design. Mentored research engineers

and scientists and acted as science lead for technical transfer in matrix organization. Supervised graduate students over summer internships.

Established new paradigms in data mining from query logs and machine-learned query rewriting in sponsored search which led to significant revenue-per-search increases and became a standard platform in Yahoo! Search Marketing.

Senior Research Scientist May 2002–January, 2004
Overture Services Pasadena, CA
Conducted research with applications to sponsored search. Mentored research engineers and acted as science lead for technical transfer. Mined query logs for term relatedness to increase sponsored search coverage and relevance, leading to significant revenue-per-search increases. Overture was acquired by Yahoo! in January 2004.

Graduate Research Assistant September 1995–September 2001
Carnegie Mellon University Pittsburgh, PA
Conducted research in web data mining, machine learning, information retrieval and spoken retrieval and translation systems. Advisor: Professor Tom Mitchell

Visiting Researcher February 1998–July 1998
Bar Ilan University Ramat Gan, ISRAEL
Conducted research on active learning under Ido Dagan.

Summer Intern June 1997–August 1997
Digital Cambridge Research Labs Cambridge, MA
Conducted research on speech recognition on lossily compressed speech data. Mentor: Chris Weikart.

Summer Intern May 1995–August 1995
Rank Xerox Research Centre Europe Grenoble, FRANCE
Modified text summarization algorithms from English to French and German. Mentor: Greg Grefenstette.

Research Engineer January 1995–May 1995
Microsoft Research Institute Sydney, AUSTRALIA
Implemented algorithms for statistical natural language processing.

SKILLS Programming Languages: Perl, C, C++, Python, Java
Tools: Unix, R, hadoop, bash, SQL, Matlab

HUMAN Fluent German, Fluent French, Conversational Japanese, Beginning Mandarin Chinese
LANGUAGES

VOLUNTEER WORK Established and ran Math Olympiad in Elementary Schools (MOEMS) club at Dr. Martin Luther King Jr. School, Cambridge MA. 2016-2017

PUBLICATIONS

Journal Papers N. Fuhr, A. Gianchanou, G. Grefenstette, I. Gurevych, A. Hanselowski, K. Jarvelin, R. Jones, Y. Liu, J. Mothe, W. Nejdl, I. Peters and B. Stein, *An Information Nutrition Label for Online Documents* SIGIR Forum 2017, to appear.

E. Riloff and R. Jones *A Retrospective on Mutual Bootstrapping* AI Magazine, 2017, to appear.

R. Jones, W. V. Zhang, B. Rey, P. Jhala and E. Stipp. *Geographic Intention and Modification in Web Search*, International Journal of Geographical Information Science (IJGIS), 22 (3), pp 229-246, 2008

R. Jones and F. Diaz *Temporal Profiles of Queries*, ACM Transactions on Information Systems (TOIS), 25 (3), 2007

H. Raghavan, O. Madani and R. Jones, *Active Learning on Both Features and Documents*, Journal of Machine Learning Research (JMLR), 7(Aug):1655–1686, 2006

R. Jones, K. Bartz, P. Subasic and B. Rey. *Automatically Generating Related Queries in Japanese*, Language Resources and Evaluation (LRE) December 2006, vol 40, numbers 3-4, pp./ 219-232

R. Ghani, R. Jones and D. Mladenic *Building Minority Language Corpora by Learning to Generate Web Search Queries*, Knowledge and Information Systems (KAIS), 2003

**Refereed
Conference
Papers**

J. Jiang, A. H. Awadallah, R. Jones, U. Ozertem, I. Zitouni, R. G. Kulkarni and O. Z. Khan *Automatic Online Evaluation of Intelligent Assistants WWW '15: Proceedings of the 24th International Conference on the World Wide Web*, 2015

A. H. Awadallah, R. G. Kulkarni, U. Ozertem and R. Jones *Characterizing and Predicting Voice Query Reformulation*, CIKM 2015

M. Shokouhi, R. Jones, U. Ozertem, K. Raghunathan and F. Diaz *Mobile Query Reformulations*, SIGIR 2014

Y. Wang, R. Khardon, D. Pechyony and R. Jones, *Generalization Bounds for Online Learning Algorithms with Pairwise Loss Functions*, The Annual Conference on Learning Theory (COLT), 2012

D. Cerrato, R. Jones and A. Gupta, *Classification of proxy labeled examples for marketing segment generation* in KDD 2011

U. Ozertem, R. Jones and B. Dumoulin, *Evaluating New Search Engine Configurations with Pre-existing Judgments and Clicks* in WWW 2011

H. Feild, J. Allan, and R. Jones, *Predicting Searcher Frustration*, Proceedings of the 33rd Annual ACM SIGIR Conference (SIGIR 2010).

A. Hassan, R. Jones and K. L. Klinkner *Beyond DCG: User Behavior as a Predictor of a Successful Search* In the Proceedings of the ACM Conference on Web Search and Data Mining (WSDM 2010), 2010.

B. Piwowarski, G. Dupret and R. Jones *Mining User Web Search Activity with Layered Bayesian Networks or How to Capture a Click in its Context*. In Proceedings of WSDM 2009

C. Barr, R. Jones and M. Regelson. *The Linguistic Structure of English Web-Search Queries*. In Proceedings of EMNLP 2008: conference on Empirical Methods in Natural Language Processing, 2008.

R. Jones and K. Klinkner. *Beyond the Session Timeout: Automatic Hierarchical Segmentation of Search Topics in Query Logs*. In Proceedings of ACM 17th Conference

on Information and Knowledge Management (CIKM 2008), Napa Valley, CA, October 2008.

R. Jones, R. Kumar, B. Pang and A. Tomkins. *Vanity Fair: Privacy in Querylog Bundles*. In Proceedings of ACM 17th Conference on Information and Knowledge Management (CIKM 2008), Napa Valley, CA, October 2008.

B. Carterette and R. Jones *Evaluating Search Engines by Modeling the Relationship Between Relevance and Clicks* in Proceedings of Twenty-First Annual Conference on Neural Information Processing Systems (NIPS 2007).

J. Teevan, E. Adar, R. Jones and M. Potts, *Information Re-Retrieval: Repeat Queries in Yahoo's Logs*, In Proceedings of the 30th Annual ACM Conference on Research and Development in Information Retrieval (SIGIR '07), Amsterdam, The Netherlands, July 2007.

R. Jones, B. Rey, O. Madani and W. Greiner, *Generating Query Substitutions* The 15th International World Wide Web Conference (WWW-2006), June 2006.

H. Raghavan, O. Madani and R. Jones: *InterActive Feature Selection* IJCAI 2005

F. Diaz and R. Jones: *Using Temporal Profiles of Queries for Precision Prediction* SIGIR 2004

L. Mayfield Tomokiyo and R. Jones, *You're Not From 'Round Here, Are You? Naive Bayes Detection of Non-native Utterance Text*, NAACL 2001.

R. Ghani, R. Jones and D. Mladenic: *Using the Web to Create Minority Language Corpora*, 10th International Conference on Information and Knowledge Management (CIKM-2001)

R. Ghani, R. Jones and D. Mladenic: *On-line learning for Web query generation: finding documents matching a minority concept on the Web*, Proceedings of the The First Asia-Pacific Conference on Web Intelligence (WI-2001)

E. Riloff and R. Jones *Learning Dictionaries for Information Extraction Using Multi-level Boot-strapping* (AAAI 1999)

**Refereed
Workshop Papers
and Conference
Posters**

D. Pechyony, R. Jones and X. Li *A Joint Optimization of Incrementality and Revenue to Satisfy Both Advertiser and Publisher*, in WWW 2013

D. Pechyony, L. Shen and R. Jones *Solving Large Scale Linear SVM with Distributed Block Minimization* in NIPS 2011 Workshop on Big Learning: Algorithms, Systems, and Tools for Learning at Scale. Granada, Spain. Dec. 16 - 17, 2011

D. Metzler, R. Jones, F. Peng, and R. Zhang *Improving Search Relevance for Implicitly Temporal Queries* Poster paper at SIGIR 2009

A. Hassan, R. Jones and F. Diaz, *Geographic Features in Web Search Retrieval*, In Proceedings of GIR 2008 Fifth Workshop on Geographic Information Retrieval, 2008

R. Jones, R. Kumar, B. Pang and A. Tomkins. *I Know What You Did Last Summer: Query Logs and User Privacy*, CIKM 2007

W. V. Zhang, X. He, B. Rey and R. Jones, *Query Rewriting using Active Learning for Sponsored Search*, Poster Paper in Proceedings of the 30th Annual ACM Conference on Research and Development in Information Retrieval (SIGIR '07), Amsterdam, The Netherlands, July 2007.

J. Teevan, E. Adar, R. Jones and M. Potts. *History Repeats Itself: Repeat Queries in Yahoo's Logs* Poster Paper In Proceedings of the 29th Annual ACM Conference on Research and Development in Information Retrieval (SIGIR '06), Seattle, WA, August 2006.

W. V. Zhang, B. Rey, E. Stipp and R. Jones: *Geomodification in Query Rewriting* Workshop on Geographic Information Retrieval, 29th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval (SIGIR 2006), Seattle WA, August 2006

B. Liu, R. Jones and K. Klinker: *Measuring the Meaning in Time Series Clustering of Text Search Queries* To appear as poster paper in the proceedings of CIKM 2006

B. Carterette, R. Jones, W. Greiner and C. Barr *N Semantic Classes are Harder Than Two* In Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions

R. Jones, M. Regelson and W. V. Zhang *Finding Grace Hopper on the Web* Poster at Grace Hopper Celebration of Women in Computing 2006

G. Kumaran, R. Jones, O. Madani: *Biasing web search results for topic familiarity* CIKM 2005: 271-272

R. Jones: *Semi-supervised Learning on Small Worlds* Link Discovery Workshop at KDD 2004

R. Jones and D. C. Fain *Query Word Deletion Prediction* Poster at 26th Annual International ACM SIGIR Conference (SIGIR 2003)

R. Jones, R. Ghani, T. Mitchell and E. Riloff: *Active Learning with Multiple View Feature Sets* ECML 2003 Workshop on Adaptive Text Extraction and Mining

R. Jones and R. Ghani: *Automatically Building a Corpus of a Minority Language from the Web* ACL 2000 Student Research Workshop

R. Ghani, R. Jones, D. Mladenic, K. Nigam, and S. Slattery, *Datamining on Symbolic Knowledge Extracted from the Web*. In KDD 2000 Text Mining Workshop

R. Jones, A. McCallum, K. Nigam and E. Riloff, *Bootstrapping for Text Learning Tasks* in IJCAI-99 Workshop on Text Mining: Foundations, Techniques and Applications 1999

O. Glickman and R. Jones, *Examining Machine Learning for Adaptable End-to-End Information Extraction Systems*, AAAI 1999 Workshop on Machine Learning for Information Extraction

**Other
publications**

R. Jones: *Learning to Extract Entities from Labeled and Unlabeled Text* PhD dissertation Carnegie Mellon University School of Computer Science, 2005

R. Ghani and R. Jones: *Automatic Training Data Collection For Semi-Supervised*

Learning of Information Extraction Systems Accenture Technology Labs Technical Report 2002

A. G. Hauptmann, R. E. Jones, K. Seymore, M. A. Siegler, S. T. Slattery and M. J. Witbrock, *Experiments in Information Retrieval from Spoken Documents*, BNTUW-98 Proceedings of the DARPA Workshop on Broadcast News Understanding Systems, Lansdowne, VA, February, 1998.

M. A. Siegler, M. J. Witbrock, S. T. Slattery, K. Seymore, R. E. Jones, and A. G. Hauptmann, *Experiments in Spoken Document Retrieval at CMU*, TREC-6 Proceedings of the 1997 Text REtrieval Conference, Gaithersburg, MD, November 19-21, 1997.

SCIENTIFIC
LEADERSHIP

Presented tutorials on web mining, and open search tools (SIGIR 2008, SIGIR 2009, WWW 2010). Organized workshops on web search click data, lexical cohesion, and unsupervised data for machine learning and data mining (WSDM 2009, SIGIR 2005 and ICML 2003).

REVIEWING

Senior Program Committee Member
COLING, SIGIR

Program Committee Member
EMNLP, RIAO, WWW, ICML, KDD, HLT/NAACL, CIKM, ACL, AAAI

Journal Reviewer
ACM TOIS
ACM TWEB
Journal of Machine Learning Research (JMLR)
Journal of Artificial Intelligence Research (JAIR)
IEEE Transactions on Multimedia
IEEE Information Systems
Information Processing and Management (IP&M)

INTERNS

Mihajlo Grbovic, Temple University, May - Aug 2012 “Feature Selection for Display Advertising Purchase Prediction”
Now at Airbnb

Prakash Comar, Michigan State University, May - Aug 2012 “Learning Incremental Purchasers”

Bernie Wang, Tufts University - May - Aug 2011 “Online Learning for Purchase Prediction”
Now at Amazon

Na Dai, Lehigh University - May - August 2011 “Learning to Predict Purchases with Sequence Data”
Now at Google

Ahmed Hassan, U Michigan, Ann Arbor May - Aug 2009 “Predicting User Search Success”; May - Aug 2008 “Geographic Information Retrieval”
Now at Microsoft Research

Desislava Petkova, UMass Amherst, May - Aug 2007 “Query Expansion for Retrieving

Answers to Questions”
Now at Procter & Gamble

Kristina Klinkner, Carnegie Mellon University, May - Aug 2007 “Automatically Segmenting Search Sessions”; May - Aug 2006 “Classifying search queries using time series”
Now at Amazon

Ben Carterette, UMass Amherst, May - Aug 2006 “Learning to Rank from Clicks”; May - Aug 2005 “Machine Learning for Classifying Semantic Classes”
Now associate professor at U. Delaware

Bing Liu, Virginia Tech, May - Aug 2005 “Clustering time series of search queries”
Now at Google

Giridhar Kumaran, UMass Amherst, May - Aug 2004, “Classifying Documents by Topic Familiarity”
Now at Expedia

Hema Raghavan, UMass Amherst, May - Aug 2004 “Feature Selection for Document Classification”
Now at LinkedIn

Fernando Diaz, UMass Amherst May - Aug 2003 “Temporal Profiles of Queries”
Now at Spotify

PATENTS

Dmitry Pechyony, Libin Shen and Rosie Jones *Parallel training of a support vector machine (SVM) with distributed block minimization* US Patent no: 9,569,401, Issued February 2017

Henry Feild, Omer Emre Velipasaoglu, Benoit Dumoulin, Elizabeth Churchill, Rosemary Jones and Jeffrey Bardzell *Systems and methods for providing search assistance technologies based on user self-efficacy and search frustration* US Patent no: 8,903,845 Issued December 2014

Rosie Jones, Fernando Diaz and Ahmed Hassan Awadallah *System and method of geo-based prediction in search result selection* US Patent no: 8,352,466, Issued January 2013

Rosie Jones, Donald Metzler and Fuchun Peng *Identifying and expanding implicitly temporally qualified queries* US patent no: 8,156,111 Issued April 10, 2012

Rosie Jones, Giridhar Kumaran, and Omid Madani *System and method for biasing search results based on topic familiarity* US Patent No. 8,095,487 Issued January 10th, 2012

Rosie Jones, Benjamin Rey and Tina Weyand *System and method for generating substitutable queries* US Patent No. 7,962,479 Issued June 14th, 2011

Kristina Klinkner and Rosie Jones *Segmentation of search topics in query logs* US Patent No. 7,877,389 Issued January 25, 2011

Rosie Jones and Desislava Petkova *Context-sensitive query expansion* US Patent No. 7,831,588 Issued November 9, 2010

Tina Weyand, Rosie Jones and Benjamin Rey *Alternative search query processing in a term bidding system* US Patent No. 7,752,220 Issued July 6, 2010

Rosie Jones, Benjamin Rey and Marco Zaghera *System and method for generating substitutable queries on the basis of one or more features* US Patent No. 7,739,264 Issued June 15, 2010

Rosie Jones, Kevin Bartz and Benjamin Rey *System and method for identifying related queries for languages with multiple writing systems* US Patent No. 7,689,554 Issued March 30, 2010

Soren Riise, Devesh Patel, Bruce Campbell, David Richardson-Bunbury, Eugene Stipp and Rosie Jones *System and method for query expansion* US patent no: 7,685,119 Issued March 23, 2010

Tina Weyand, Rosie Jones and Benjamin Rey *System and method for determining alternate search queries* US Patent No. 7,634,462 Issued December 15, 2009

Rosie Jones and Fernando Diaz *System and method for providing temporal search results in response to a search query* US Patent No. 7,577,651 Issued August 18, 2009

Rosemary Jones and Daniel C. Fain, *System and Methods for Ranking the Relative Value of Terms in a Multi-Term Search Query Using Deletion Prediction* US Patent No. 7,406,465 Issued July 29, 2008

AWARDS AND
HONORS

Senior Member of ACM 2008 - present

Yahoo! Internal Patent Milestone Award in recognition of my 15th Yahoo! invention, February 2008

Nominated for Yahoo! Internal Superstar Award, 2006

Yahoo! Internal Roosevelt Teddy Award, August 2006

National Science Foundation LIS scholarship for multimedia research. Half-tuition and stipend award. Spring 1999, Summer 2000, Fall 2000

Canon Australia Scholarship for Excellence in Computer Science AUS\$10,000 prize awarded May 1994

1991 OTC Scholarship for Women in Electrical Engineering Awarded on the basis of academic merit

Sydney University Chancellor's Scholarship in Engineering, awarded 1990, on the basis of academic merit