

## ROSIE JONES

---

Yahoo! Inc  
4 Cambridge Center  
Cambridge, MA 02142

Phone: (617)806-8248  
Email: [rosie.jones@acm.org](mailto:rosie.jones@acm.org)  
Web: <http://www.cs.cmu.edu/~rosie/>

RESEARCH INTERESTS Information retrieval, web search, geographic information retrieval, computational advertising, Asian language processing.

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) Computer Science, 1994

RESEARCH EXPERIENCE Senior Research Scientist, Information Retrieval May 2002–Present  
Yahoo / Overture Burbank / Pasadena, CA  
Conduct research in information retrieval with applications to web and sponsored search. Use machine learning and statistical analysis of data to design and run experiments, test hypotheses leading to algorithm design. Mentor research engineers and act as science lead for technical transfer.

Graduate Research Assistant September 1997 - September 2001  
Carnegie Mellon University Pittsburgh, PA  
A research member in the Web->KB research project, led by Prof. Tom Mitchell. Performed research in information extraction. Assisted in building a web-crawling agent that extracts information from the Web into a propositional knowledge base, using a variety of learned text classifiers and information extractors.

Visiting Researcher Feb 1998 - July 1998  
Bar Ilan University Ramat Gan, ISRAEL  
Conduct research on active learning under Ido Dagan.

Graduate Research Assistant June 1996 - June 1997  
Carnegie Mellon University Pittsburgh, PA  
A research member in the Listen and Learn research project, led by Prof. Tom Mitchell. Performed research on finding speech confidence measures for machine learning algorithms for classification of radio and television news.

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

Graduate Research Assistant Sept 1995 - May 1996  
Carnegie Mellon University Pittsburgh, PA  
JANUS Project: English grammar writing, rescoring lattices using information from

link grammar parser and out-of-domain vocabulary detection in the context of machine translation of spoken dialogs.

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

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

#### JOURNAL PAPERS

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. Mladenovic *Building Minority Language Corpora by Learning to Generate Web Search Queries* Knowledge and Information Systems (KAIS), 2003

#### REFEREED CONFERENCE PAPERS

B. Piwowarski, G. Dupret and Rosie 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. 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.

Invited  
Presentations

*"I know what you did last summer": query logs and user privacy* UCLA, June 2008

*Geographic Intention and Modification in Web Search*, UCSD, May 2008

*Geographic Intention and Modification in Web Search*, UMass Amherst, March 2008

*Learning from User Query Reformulations*, UIUC October 2007

*Learning from User Query Reformulations*, Microsoft Research, August 2006

*Temporal Profiles of Queries*, Carnegie Mellon University, October 2005

*What's the Most Frequent Word in English and Other Natural Language Fun with Perl*  
yapc Perl conference, Pittsburgh, 1999

## Professional Activities

Tutorials:

SIGIR 2009 tutorial on IR Prototypes and Web Search Hacks with Open Source Tools,  
Rosie Jones and Vik Singh

SIGIR 2008 tutorial on web mining for search, Ricardo Baeza-Yates and Rosie Jones

Workshop Organization:

WSDM 2009: Workshop on *Web Search Click Data (WSCD 09)*

SIGIR 2005 workshop: *Methodologies and Evaluation of Lexical Cohesion Techniques  
in Real-world Applications (Beyond Bag of Words)*

ICML 2003 Workshop on *The Continuum from Labeled to Unlabeled Data in Machine  
Learning and Data Mining*

Yahoo! Research Summer Intern Poster Session: instigated and coordinated 2004, 2005

Yahoo! Research Workshop on Social Networks, June 2004

Reviewing: SIGIR 2007, 2008, 2009: Senior Program Committee member (equivalent  
to area chair)

SIGIR 2005,2006

WWW 2006, 2008, 2009

ICML 2008

KDD 2008

HLT/NAACL 2007

CIKM 2006

ACL-COLING 2006

AAAI 2006

ACM TOIS

ACM TWEB

Journal of Machine Learning Research (JMLR)

Journal of Artificial Intelligence Research (JAIR)

IEEE Transactions on Multimedia

IEEE Information Systems HLT/NAACL 2004

Information Processing and Management (IP&M)

## INTERNS

Ahmed Hassan, U Michigan, Ann Arbor May - Aug 2009 "Predicting User Search Suc-  
cess"

Ahmed Hassan, U Michigan, Ann Arbor May - Aug 2008 "Geographic Information  
Retrieval"

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

Answers to Questions”

Kristina Klinkner, Carnegie Mellon University, May - Aug 2007 “Automatically Segmenting Search Sessions”

Ben Carterette, UMass Amherst, May - Aug 2006 “Learning to Rank from Clicks”

Kristina Klinkner, Carnegie Mellon University: May - Aug 2006 “Classifying search queries using time series”

Ben Carterette, UMass Amherst: May - Aug 2005 “Machine Learning for Classifying Semantic Classes”

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

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

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

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

## PATENTS

### Patents Awarded

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

### Patents Filed

20080154856 System and method for query expansion

20080140699 System and Method for Generating Substitutable Queries

20080114721 System and method for generating substitutable queries on the basis of one or more features

20080077588 Identifying and Measuring Related Queries

20070203894 System and method for identifying related queries for languages with multiple writing systems

20070038621 System and method for determining alternate search queries

20070038602 Alternative search query processing in a term bidding system

System and method for biasing search results based on topic familiarity

System and method for providing interactive feature selection for training a document classification system

## AWARDS AND HONORS

*Senior Member of ACM* 2008 - present

*Yahoo! Internal Patent Milestone Award* in recognition of 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

#### SKILLS

Programming Languages: Perl, C, C++

*Los Angeles Perl Monger's Beginner's Perl Golf Challenge Winner* April 2002

Human Languages: Fluent German, Fluent French, Advanced Conversational Japanese (passed JLPT level 4), Elementary Hebrew (completed ulpan kita aleph)

Extensive overseas working and foreign language experience in France (total 14 months), Japan (3 months), Germany (7 months), Israel (5 months), USA (10 years)