

Jaime Arguello

Curriculum Vitae

EDUCATION

Ph.D., Language Technologies Advisor: Dr. Jamie Callan Language Technologies Institute, School of Computer Science Carnegie Mellon University	2006-Present
M.S., Language Technologies Advisor: Dr. Carolyn Rosé Language Technologies Institute, School of Computer Science Carnegie Mellon University	2005-2006
M.S., Information Systems Management School of Information Systems and Management Heinz College Carnegie Mellon University	2003-2005
B.S., Electrical Engineering Minor, Economics Washington University (St. Louis, MO)	1996-2000

RESEARCH EXPERIENCE

Language Technologies Institute, Carnegie Mellon University 2006-Present
Pittsburgh, PA

My dissertation focuses on *aggregated search*, the task of incorporating results from multiple specialized search engines, known as *verticals*, into Web search results. This includes methods for *vertical selection*—the task of deciding which verticals to present—and methods for *results merging*—the task of deciding where in the Web results to embed the vertical results. My methods are informed by prior work in federated search, but use machine learning to harness diverse sources of evidence. My most recent work focuses on a metric-based evaluation methodology for aggregated search.

Prior to this, I conducted research in various areas of information retrieval. I worked on *blog retrieval*—the task of retrieving blogs with a recurring interest in the topic of the query. Also, within the eRuleMaking project, I worked on semi-supervised methods for identifying stakeholder communities and recognizing evidence-citation in public comment corpora (emails sent by the public to a government agency that is proposing new regulation).

During my masters, I worked with Dr. Carolyn Rosé on developing tools for authoring computer dialogue systems from transcribed task-oriented human-human conversations. My focus was on topic-segmentation of transcribed dialogue and on developing visualizations and interactions for organizing dialogue-segments into thematic clusters.

Yahoo! Laboratories Summer 2008 and Fall 2009
Montreal, QC

I did two internships at Yahoo! Labs. During the first, I worked on a machine learning approach to vertical selection. During the second, I worked on domain adaptation for vertical selection. Our goal was to learn a predictive model for one vertical using training data from another.

TEACHING EXPERIENCE

Search Engines and Web Mining—Guest Lecturer Fall 2010
Carnegie Mellon University
Professors: Jamie Callan and Yiming Yang
Prepared slides and conducted two guest lectures on federated search for an undergraduate information retrieval course.

Information Retrieval—Teaching Assistant Spring 2010
Carnegie Mellon University
Professors: Jamie Callan and Yiming Yang
Prepared slides and conducted two guest lectures on federated search for a graduate information retrieval course. Designed final exam questions, graded homework, and held office hours.

PUBLICATIONS

Journal Articles

J. Arguello, J. Callan, and Stuart Shulman. Recognizing Citations in Public Comments. In John Wilkerson and Claire Cardie, editors, *Special Issue on Text Annotation for Political Science of the Journal of Information Technology and Politics*. 5(1). 2008.

Refereed Papers in Archival Conference Proceedings

J. Arguello, F. Diaz, J. Callan, and B. Carterette. A Methodology for Evaluating Aggregated Search Results. To appear in the *European Conference in Information Retrieval*. Dublin. 2011. [acceptance rate: 20%]

J. Arguello, F. Diaz, and J.-F. Paiement. Vertical Selection in the Presence of Unlabeled Verticals. In *SIGIR '10: Proceedings of the 33rd International ACM Conference on Research and Development in Information Retrieval*. Geneva. 2010. [acceptance rate: 17%]

J. Arguello, J. Callan, and F. Diaz. Classification-based Resource Selection. In *CIKM '09: Proceedings of the 18th ACM Conference on Information and Knowledge Management*. Hong Kong. 2009. [acceptance rate: 15%]

J. Arguello, F. Diaz, J. Callan, and J.-F. Crespo. Sources of Evidence for Vertical Selection. In *SIGIR '09: Proceedings of the 32nd International ACM Conference on Research and Development in Information Retrieval*. Boston. 2009. [acceptance rate: 16%] (**Best Paper Award**)

F. Diaz and **J. Arguello**. Adaptation of Offline Vertical Selection Predictions in the Presence of User Feedback. In *SIGIR '09: Proceedings of the 32nd International ACM Conference on Research and Development in Information Retrieval*. Boston. 2009. [acceptance rate: 16%]

J. Elsas, **J. Arguello**, J. Callan, and J. Carbonell. Retrieval and Feedback Models for Blog Feed Search. In *SIGIR '08: Proceedings of the 31st International ACM Conference on Research and Development in Information Retrieval*. Singapore. 2008. [acceptance rate: 17%]

J. Arguello, J. Elsas, J. Callan, and J. Carbonell. Document Representation and Query Expansion Models for Blog Recommendation. In *ICWSM '08: Proceedings of the 2nd International IEEE Conference in Weblogs and Social Media*. Seattle. 2008. [acceptance rate: 25%] (**Best Paper Nomination**)

J. Arguello and J. Callan. A Bootstrapping Approach for Identifying Stakeholders in Public-Comment Corpora. In *DG.O '07: Proceedings of the 8th International Conference in Digital Government Research*. Philadelphia. 2007.

J. Arguello and C. Rosé. Museli: A Multi-source Evidence Integration Approach to Topic-Segmentation of Dialogue. In *HLT-NAACL '06: Proceedings of the Human Language Technology Conference of the NAACL*. New York. 2006. (short-paper) [acceptance rate: 24%]

J. Arguello and C. Rosé. InfoMagnets: Making Sense of Corpus Data. In *HLT-NAACL '06: Proceedings of the Human Language Technology Conference of the NAACL*. New York. 2006. (system demonstration) [acceptance rate: 24%]

J. Arguello, B. Butler, L. Joyce, R. Kraut, K. Ling, C. Rosé, and X. Wang. Talk to Me: Foundations of Successful Individual-Group Interactions in Online Communities. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. Montreal. 2006. [acceptance rate: 20%]

C. Rosé, C. Pai, and **J. Arguello**. Enabling Non-Linguists to Author Advanced Conversational Interfaces Easily. In *FLAIRS '05: Proceedings of the 18th International Florida Artificial Intelligence Research Society Conference*. Clear Water Beach. 2005

Refereed Workshop Papers

J. Arguello and C. Rosé. Topic Segmentation of Dialogue. In *Proceedings of the HLT-NAACL 2006 Workshop on Analyzing Conversations in Text and Speech*. New York. 2006.

G. Gweon, **J. Arguello**, C. Pai, R. Carey, Z. Zaiss, and C. Rosé. Towards a Prototyping Tool for Behavior-Oriented Authoring of Conversational Agents for Educational Applications. In *Proceedings of the ACL 2005 Workshop for Building Educational Application using NLP*. Ann Arbor. 2005.

Other Publications

J. Arguello, J. Elsas, C. Yoo, J. Callan, and J. Carbonell. Document and Query Expansion Models for Blog Distillation. In *TREC '08: Proceedings of the 2008 Text REtrieval Conference*. National Institute of Standards and Technology. 2008.

J. Elsas, **J. Arguello**, J. Callan, and J. Carbonell. Retrieval and Feedback Models for Blog Distillation. In *TREC '07: Proceedings of the 2007 Text REtrieval Conference*. National Institute of Standards and Technology. 2007. **(Best Performing System)**

PRESENTATIONS

Vertical Selection in the Presence of Unlabeled Verticals, SIGIR 2010, Geneva

Vertical Selection in the Presence of Unlabeled Verticals, Yahoo! Labs, Montreal, 2010

Classification-based Resource Selection, CIKM 2009, Hong Kong

Sources of Evidence for Vertical Selection, SIGIR 2009, Boston

Sources of Evidence for Vertical Selection, Yahoo! Labs, Montreal, 2009

Document Representation and Query Expansion Models for Blog Recommendation, ICWSM 2008, Seattle

A Bootstrapping Approach for Identifying Stakeholders in Public Comment Corpora, DG.O 2007, Philadelphia

InfoMagnets: Making Sense of Corpus Data. Demo Plenary, HLT-NAACL 2006, New York

Museli: A Multi-source Evidence Integration Approach to Topic-Segmentation of Dialogue, HLT-NAACL 2006, New York

Topic Segmentation of Dialogue, HLT-NAACL 2006, New York

Topic Segmentation of Dialogue, LTI Symposium, CMU, August 2006, Pittsburgh

AWARDS

Best Paper Award, SIGIR 2009, for *Sources of Evidence for Vertical Selection*
Yahoo! Key Scientific Challenges Award, 2009
Best Paper Nomination, ICWSM 2008, for *Document Representation and Query Expansion Models for Blog Recommendation*

SERVICE

Academic

Reviewer for ECIR 2011
Reviewer for SIGIR 2010
Reviewer for COLING 2010
Reviewer for JASIST, 2010
Reviewer for Data Challenge Workshop, ICWSM 2009
Reviewer for ECIR 2008
Reviewer for WWW 2008
Program Committee Member, Program Chair of Student Session, DG.O 2008
Co-organizer of Information Retrieval Seminar, Carnegie Mellon, 2007-Present

Local Community

The Pittsburgh Sprout Fund September 2009 and 2010
I volunteered in a fundraising event to support artistic development in Pittsburgh.

Pittsburgh Social Venture Partners May 2007- May 2008
The PSVP advises and provides technical support to other non-profits that serve at-risk youth. My one-year membership was granted by being awarded the Recker Fellowship.

RELEVANT GRADUATE COURSEWORK

Advanced Information Retrieval Seminar	Grammars and Lexicons
Advanced Topics In User Interfaces	Information Retrieval
Algorithms for NLP	Language and Statistics I & II
Conversational Interfaces	Machine Learning
Digital Libraries	Software Architectures for User Interfaces
Experimental Information Retrieval	Text Data Mining

CONTACT INFORMATION

Jaime Arguello
5515 Gates-Hillman Complex
5000 Forbes Ave.
Pittsburgh, PA, USA, 15213
T (630) 781-3898
F (412) 268-6298

jaime@cs.cmu.edu
<http://www.cs.cmu.edu/~jaime>

REFERENCES

Jamie Callan (academic advisor)
Language Technologies Institute, School of Computer Science
Carnegie Mellon University
callan@cs.cmu.edu

Jaime Carbonell
Language Technologies Institute, School of Computer Science
Carnegie Mellon University
jgc@cs.cmu.edu

Carolyn Rosé
Language Technologies Institute, School of Computer Science
Carnegie Mellon University
cprose@cs.cmu.edu

Fernando Diaz
Yahoo! Research
diazf@yahoo-inc.com