

Language Technologies Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15207
United States of America

home: +1 412 421 2795
work: +1 412 268 5606
mobile: +1 412 901 9886
email: pto@cs.cmu.edu
<http://www.cs.cmu.edu/~pto>

Paul Ogilvie

Research Interests

Information retrieval, retrieval of structured documents (including XML), web document retrieval, statistical evaluation methodology in information retrieval, retrieval support for natural language applications

Education

- 2000-Present • **Ph.D.** in *Language and Information Technologies* (expected August 2007). Carnegie Mellon University, Language Technologies Institute, School of Computer Science. Dissertation title: Generative Probabilistic Models for Retrieval of Documents with Structure and Annotations. Advisor: Prof. Jamie Callan
- 2000-2003 • **Master of Science** in *Language and Information Technologies*. Carnegie Mellon University, Language Technologies Institute, School of Computer Science
- 1996-2000 • **Bachelor of Science** in *Computer Science and Mathematics, summa cum laude*. University of Massachusetts at Amherst

Professional Experience

- 2000-Present • **Research Assistant**, *Language Technologies Institute*, Carnegie Mellon University. Researched web document retrieval, XML element retrieval, and retrieval leveraging linguistic annotations. Member of Lemur project steering committee and contributor to the Lemur Toolkit for Information Retrieval. Researched evaluation methodology for XML retrieval while a member of the Metrics Group for the Initiative for the Evaluation of XML Retrieval.
- Summer 2006 • **Graduate Intern**, *IBM Almaden Research Center*. Developed solutions for information extraction. Assisted with design of an evaluation for a user tool to create rule-based annotators. Built a web-based hand annotation tool communicating with a remote database to allow creation of gold-standard assessments.
- June 2003 • **Researcher**, *Reliable Information Access (RIA) Workshop*, Northeast Regional Research Center, MITRE, Boston, Massachusetts. Researched and performed failure analysis of several pseudo-relevance feedback techniques of several major information retrieval systems.
- Summer 2000 • **Intern**, *AOL Search Group*, Dulles, Virginia. Integrated a part-of-speech tagger into the AOL search engine text processing architecture.
- Jan. 1998 - May 2000 • **Senior Undergraduate Research Programmer**, *Center for Intelligent Information Retrieval*, University of Massachusetts at Amherst. Designed and developed visualization interfaces for the InQuery search engine. These visualizations depicted query-document and document-document similarity in an interactive environment.

Professional Experience (continued)

- Summer 1999
- **Researcher**, *National Science Foundation Research Experience for Undergraduates*, University of Massachusetts at Amherst. Developed a rule-based extractor for acronyms and their expansions. Built a large web-accessible database of acronym definitions. Constructed a testbed and evaluated the accuracy of extracted definitions.

Teaching Experience

- Nov. 2006
- **Invited Lecture**, *XML Retrieval*. Presented to the graduate level course “Principles of Information Retrieval” (IMT 542) offered by the Information School at the University of Washington, Seattle, Washington.
- Aug. 2006
- **Tutorial Presenter**, *Using the Lemur Toolkit for Information Retrieval*. ACM SIGIR International Conference on Research and Development in Information Retrieval, Seattle, Washington. There were 28 attendees to this 1/2 day hands-on tutorial presented with Trevor Strohman.
- Fall 2005
- **Teaching Assistant**, *Algorithms for Natural Language Processing (11-711)*, Language Technologies Institute, Carnegie Mellon University: presented weekly recitation sections, helped design homework assignments, held weekly office hours, and graded student work. This graduate level course covered formal language theory, part-of-speech tagging, grammar and statistical parsing, and other related technologies.
- Spring 2002
- **Teaching Assistant**, *Information Retrieval (11-741)*, Language Technologies Institute, Carnegie Mellon University: assisted in design of course assignments, presented one lecture, held weekly office hours, and graded student work. This graduate level course covered principles of search engines, text classification and filtering algorithms, and evaluation of information retrieval systems. I also presented occasional lectures for this course in other years.
- Spring 2002
- **Teaching Assistant**, *Organizations and Strategic Information Processing (95-784)*, Heinz School of Public Policy and Management, Carnegie Mellon University: assisted in design of course assignments, presented one lecture, held weekly office hours, and graded student work. This graduate level course covered principles of search engines and databases, as well as information and data mining techniques. I also presented occasional lectures for this course in other years.
- Spring 2000
- **Teaching Assistant and Grader**, *Operating Systems (CMPSCI 377)*, Computer Science Department, University of Massachusetts at Amherst: designed homework assignments and graded student work. This was an undergraduate level course.
- Spring 1999

Honors and Awards

- 2004-2005
- Northrup Grumman Fellowship
- 2004
- ACM SIGIR Doctoral Consortium Award
- 1996-2000
- ACORN Scholarship Recipient

Professional Service

- **Metrics group member**, Initiative for the Evaluation of XML Retrieval (INEX): 2005, 2006
- **Founder** of the Information Retrieval Discussion Series at Carnegie Mellon University. Obtained sponsorship from Yahoo!.
- **System administrator**, ACM SIGIR conference online paper review system: 2003

Professional Service (continued)

Journal Reviewing

- ACM Transactions on Information Systems
- Information Processing and Management
- Journal of Digital Libraries
- Kluwer (Springer) Information Retrieval

Conference Reviewing

- ACM SIGIR International Conference on Research and Development in Information Retrieval: Papers 2006, 2007, Posters 2005, 2006, Demos 2007
- String Processing and Information Retrieval (SPIRE): 2006
- Human Language Technology Conference - North American chapter of the Association for Computational Linguistics Annual Meeting (HLT-NAACL): Student Workshop 2003, 2004

Invited Talks

- Dec. 2006 • **Bridging Search Engines and Natural Language Applications: Preliminary Experiments in Semantic Predicate Retrieval.** Yahoo! Barcelona Research Lab and Universitat Pompeu Fabra, Barcelona, Spain.
- Nov. 2006 • **INEX Report-Out.** The Text Retrieval Conference. National Institute of Standards and Technology, Gaithersburg, Maryland.
- Dec. 2004 • **A Hierarchical Generative Language Model for XML Element Retrieval.** Center for Intelligent Information Retrieval, University of Massachusetts at Amherst.
- June 2004 • **Generative Probabilistic Models for Retrieval of Documents with Structure and Annotations.** University of Amsterdam, the Netherlands.

Publications

Conference Papers

- 2007 • Ogilvie, Paul, Matthew Bilotti, Jamie Callan, and Eric Nyberg. *Structured Retrieval for Question Answering.* To be presented at the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007).
- 2006 • Ogilvie, Paul and Mounia Lalmas. *Investigating the Exhaustivity Dimension in Content-Oriented XML Element Retrieval Evaluation.* In the Proceedings of the Fifteenth International Conference on Information Knowledge Management (CIKM 2006).
- 2005 • Ogilvie, Paul and Jamie Callan. *Experiments with Language Models for Known-Item Finding of E-mail Messages.* In the Proceedings of the Fourteenth Text Retrieval Conference (TREC-14).
- 2004 • Collins-Thompson, Kevyn, Paul Ogilvie, and Jamie Callan. *Initial Results with Structured Queries and Language Models on Half a Terabyte of Text.* In the Proceedings of the Thirteenth Text Retrieval Conference (TREC-13).
- 2003 • Ogilvie, Paul and Jamie Callan. *Combining Document Representations for Known Item Search.* In the Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2003), pp. 143-150.
- Ogilvie, Paul and Jamie Callan. *Combining Structural Information and the Use of Priors in Mixed Named-Page and Homepage Finding.* In the Proceedings of the Twelfth Text Retrieval Conference (TREC-12).
- 2002 • Si, Luo, Rong Jin, Jamie Callan, and Paul Ogilvie. *Language Model Framework for Resource Selection and Results Merging.* In the Proceedings of the Eleventh International Conference on Information and Knowledge Management (CIKM 2002), pp. 391-397.

Publications (continued)

- Collins-Thomson, Kevyn, Paul Ogilvie, Yi Zhang, and Jamie Callan. *Information Filtering, Novelty Detection, and Named-Page Finding*. In the Proceedings of the Eleventh Text Retrieval Conference (TREC-11), pp. 107-118.
- 2001 • Ogilvie, Paul and Jamie Callan. *The Effectiveness of Query Expansion for Distributed Information Retrieval*. In the Proceedings of the Tenth International Conference on Information Knowledge Management (CIKM 2001), pp. 183-190.
- Ogilvie, Paul and Jamie Callan. *Experiments Using the Lemur Toolkit*. In the Proceedings of the Tenth Text Retrieval Conference, TREC 2001. NIST Special Publication 500-250, pp. 103-108.
- 2000 • Lavrenko, Victor, Matt Schmill, Dawn Lawrie, Paul Ogilvie, David Jensen, and James Allan. *Language Models for Financial News Recommendation*. In the Proceedings of the Workshop on Text Mining at the Ninth International Conference on Information Knowledge Management (CIKM 2000), pp. 389-396.
- Larkey, Leah, Paul Ogilvie, M. Andrew Price, and Brenden Tamilio. *Acrophile: An Automated Acronym Extractor and Server*. In the Proceedings of the Fifth ACM Conference on Digital Libraries (DL '00), pp. 205-214.

Workshop Papers

- 2005 • Ogilvie, Paul and Jamie Callan. *Parameter Estimation for a Simple Hierarchical Generative Model for XML Retrieval*. In the Proceedings of the Initiative for the Evaluation of XML Retrieval Workshop (INEX 2005), pp. 211-224.
- 2004 • Ogilvie, Paul and Jamie Callan. *Hierarchical Language Models for Retrieval of XML Components*. In the Proceedings of the Initiative for the Evaluation of XML Retrieval Workshop (INEX 2004).
- Ogilvie, Paul. *Retrieval Using Structure for Question Answering*. In the Proceedings of the First Twente Data Management Workshop (TDM'04).
- 2003 • Ogilvie, Paul and Jamie Callan. *Using Language Models for Flat Text Queries in XML Retrieval*. In the Proceedings of the Initiative for the Evaluation of XML Retrieval Workshop (INEX 2003).
- 2002 • Ogilvie, Paul and Jamie Callan. *Language Models and Structured Document Retrieval*. In the Proceedings of the Initiative for the Evaluation of XML Retrieval Workshop (INEX 2002).
- 2000 • Lavrenko, Victor, Matt Schmill, Dawn Lawrie, Paul Ogilvie, David Jensen, and James Allan. *Mining of Concurrent Text and Time Series*. In the Workshop Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2000), pp. 37-44.

Other Publications

- 2004 • Ogilvie, Paul. *Understanding Combination of Evidence using Generative Probabilistic Models for Information Retrieval*. Prepared for the Doctoral Consortium at the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2004). *A winner of the Doctoral Consortium Award (also awarded to Jochen Leidner)*.
- 2000 • Ogilvie, Paul. *Extracting and Using Relationships Found in Text for Topic Tracking*. Undergraduate Honors Thesis, CIIR Technical Report IR-209.