

Barry Schiffman

barry@textscope.com

PROFESSIONAL HISTORY

Consulting

January 2009 to present

Independent consultant focusing on projects with an artificial intelligence, machine learning or natural language processing component. Currently developing software and techniques to enhance intelligent search and information extraction from text.

Columbia University

January 1998 to December 2008

Postdoc Research Scientist, Natural Language Processing Group, Columbia University, as an IC Postdoctoral Fellow. My project focused on automatically obtaining semantic and ontological information from text. I also received funding from the DARPA GALE project for work in the distillation part of the program. This involved multimodal, multilingual question-answering. I wrote modules to handle several types of questions, and also worked closely with the P.I. and with our group's programmer to co-ordinate our system, and to communicate with our partners at SRI, NYU and UMass. September 2006 to the present.

Before this, Ph.D. candidate and graduate research assistant in Computer Science, specializing in Natural Language Processing, pursuing original research with the N.L.P. Research Group. In the spring semester in 1998, taught a one-credit course on the C programming language, and in the fall semester in 1998 was the T.A. for the N.L.P. course. In summer 2002, taught Introduction to Computer Science in Java. Research involves investigation of computational linguistics problems, writing programs to perform automated analysis of texts, and collaborating with other researchers. Projects have included pronomial resolution, question-answering, information extraction, multi-document summarization, and the determination of new information.

Salary: \$5,400 / month

P.I.: Prof. Kathy McKeown

--

Mitre Corporation

November 2000 to December 2002

Consultant on various NLP projects, part-time. Worked mostly on two projects: Biogen, which extracted and presented short descriptions of people mentioned in a large corpus;

and research into event-time labeling.

Salary: \$250 / day

Supervisor: Inderjeet Mani, senior researcher.

--

The New York Times

February 1988 to January 1998

Copy editor on the National Desk, which handles articles covering national politics, and breaking news across the country, including crime, disasters, etc. as well as articles on subjects like science, education and the law, and on the Business Day Desk. Job involved rewriting and editing articles for style and content, and writing headlines.

Salary: \$80,000 / year

Supervisor: Dean Bacquet, National Editor.

--

The Wall Street Journal

July 1984 to February 1988

News editor, supervising the editing of political, regulatory, economics and banking stories, overseeing the work of 10 copy editors; previously, copy editor, rewriting, editing and headline writing.

--

The Hartford Courant

January 1971 to July 1984

Reporter and editor in various jobs for the daily and Sunday papers.

--

SKILLS

Programming in Java, Python, Perl, C++, C (note, however, good programming is language-independent). Knowledge of shell-scripting, CGI and Javascript. Familiar with maintenance and administration of Linux systems, including CentOS, Red Hat and Ubuntu. Also familiar with the Mac OS. Linguistics, computational linguistics, and some knowledge of French. Background in writing and editing of news.

-

EDUCATION

Columbia University

PhD in Computer Science, dissertation "Learning to Identify New Information" was successfully defended on May 11, 2005

MS in January 2000, in computer science. Grade-Point-Average: 4.044.

City University of New York

BA in May 1996, in computer science, through the CUNY BA program at Hunter College, with honors. Grade-Point-Average: 4.0.

Began the M.S. program at Columbia in September 1996 as a part-time student, then was accepted as a full-time PhD student.

New York University

BA in January 1969, history major; transferred from Rutgers University.

PUBLICATIONS:

"A New Framework for Unsupervised Semantic Discovery", Barry Schiffman, Columbia University Technical Report, CUCS-039-07.

"Question Answering Using Integrated Information Retrieval and Information Extraction", Barry Schiffman et al., HLT/NAACL, 2007.

"Context and Learning in Novelty Detection", Barry Schiffman and Kathleen R. McKeown, HLT/EMNLP 2005.

"Columbia University in the Novelty Track at TREC 2004", Barry Schiffman and Kathleen R. McKeown, Text Retrieval Conference 2004

"Machine Learning and Text Segmentation in Novelty Detection", Barry Schiffman and Kathleen R. McKeown, Columbia University Technical Report, CUCS-036-04, 2004

"An Investigation Into the Detection of New Information", Barry Schiffman and Kathleen R. McKeown, Columbia University Technical Report, CUCS-035-04, 2004

"Inferring Temporal Ordering of Events in News". I. Mani, B. Schiffman, J. Zhang, In Proceedings of the Human Language Technology Conference (HLT-NAACL, 2003).

- ``Experiments in Novelty Detection at Columbia University'', Barry Schiffman, in Proceedings of the Text Retrieval Conference, 2002
- ``Building a Resource for Evaluating the Importance of Sentences'', Barry Schiffman, In Proceedings of the Language Resources and Evaluation Conference, 2002
- ``Experiments in Multidocument Summarization'', Barry Schiffman, Ani Nenkova and Kathleen McKeown in Proceedings of the Human Language Technology Conference, 2002
- ``Tracking and Summarizing News on a Daily Basis with Columbia's Newsblaster'', Kathleen R. McKeown and Regina Barzilay and David Evans and Vasileios Hatzivassiloglou and Judith L. Klavans and Carl Sable and Barry Schiffman and Sergey Sigelman, in Proceedings of Human Language Technologies Conference, 2002
- ``The Columbia Multi-Document Summarizer for DUC 2002'', Kathleen McKeown, David Evans, Ani Nenkova, Regina Barzilay, Vasileios Hatzivassiloglou, Barry Schiffman, Sasha Blair-Goldensohn, Judith Klavans and Sergey Sigelman in Proceedings of the Document Understanding Conference, 2002
- ``Producing Biographical Summaries: Combining Linguistic Knowledge with Corpus Statistics'', Barry Schiffman, Inderjeet Mani and Kristian J. Concepcion, in Proceedings European Association for Computational Linguistics 2001.
- ``Experiments in Automated Lexicon Building'', Barry Schiffman and Kathleen McKeown, in Proceedings of COLING, 2000