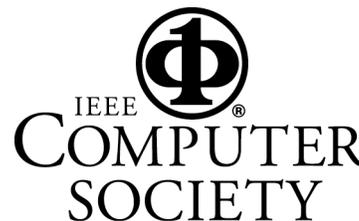


Modeling Science: Topic Models of Scientific Journals and Other Large Document Collections



David Blei

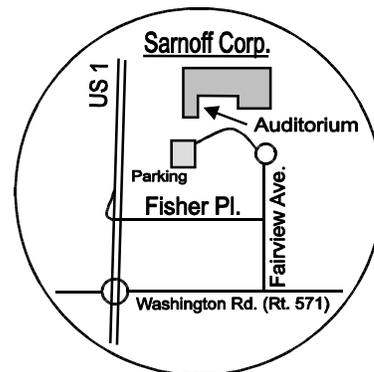
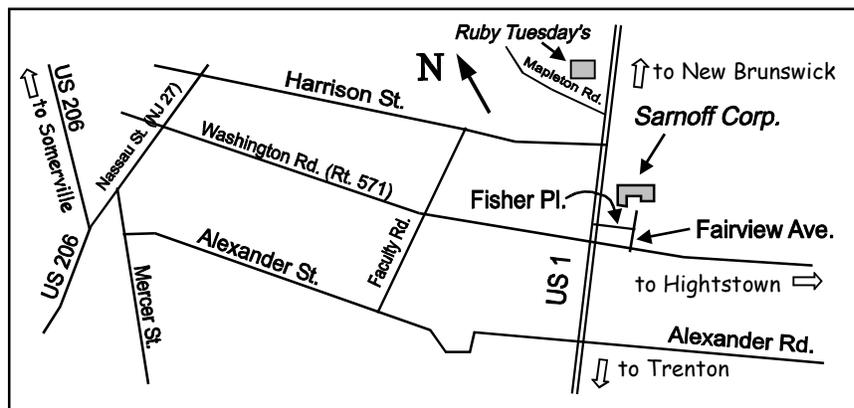
As our collective knowledge continues to be digitized and stored in on-line documents, we need new tools for organizing and annotating them. This talk will present topic models, a suite of new machine learning algorithms that examine original on-line texts to provide new methods of navigating digitized information. With topic models, we can search and explore a collection of documents based on the themes that run through it. We can “zoom in” and “zoom out” to find specific or broader themes; we can look at how those themes changed through time; we can see how themes are connected to each other. Topic models enable us to organize and summarize electronic archives at a scale that is impossible by human annotation.

David Blei is an assistant professor in the Computer Science department at Princeton University. He received his Ph.D. in 2004 from U.C. Berkeley and was a postdoctoral researcher in the Department of Machine Learning at Carnegie Mellon University. His research interests include graphical models, approximate posterior inference, and nonparametric Bayesian statistics. He focuses on applications to information retrieval and natural language processing.

Date: Thursday, January 15, 2009, 8:00 pm.
(Refreshments and networking at 7:30 pm.)
Place: Sarnoff Corp., Routes 1 and 571, Princeton, NJ
Information: Dennis Mancl (908) 582-7086, Jan Buzydlowski (610) 902-8343
On-line info: <http://www.acm.org/chapters/princetonacm>

All ACM / IEEE-CS meetings are open to the public. Students and their parents are welcome. There is no admission charge, and refreshments are served.

A pre-meeting dinner with the speaker is held at 6:00 p.m. at **Ruby Tuesday's Restaurant** on Route 1. Please send email to princetonacm@acm.org in advance if you plan to attend the dinner.



Princeton ACM calendar

Feb. 19

Social Networking: For Good and For Profit,
Peter Shankman, The Geek Factory, Inc.

Mar. 19

Reliable Extraction and Meaningful
Inference from Nonstructured Data, R.
Bharat Rao, Siemens Medical Solutions