

# Joshua Blumenstock

102 South Hall  
University of California, Berkeley  
Berkeley, CA 94720-4600

jblumenstock@berkeley.edu  
<http://www.jblumenstock.com>

## Education

- **University of California at Berkeley.** September 2006-*present*  
Ph.D. candidate, School of Information  
M.A. candidate, Department of Economics  
Areas of Concentration: Economic development, Machine learning, and Development theory.
- **Wesleyan University.** 1998-2003  
B.A. Physics and Computer Science, *Phi Beta Kappa*  
*High departmental honors in Physics*  
*High departmental honors in Computer Science*  
Honors Thesis: “Mining Clusters for Knowledge: Finding Algorithm-Independent Groups in Microarray Data”

## Honors, Scholarships, and Fellowships

2006-2010 Berkeley Graduate Fellowship, University of California at Berkeley  
2009 ICT for Agriculture Grant, Bill and Melinda Gates Foundation  
2009 Cultural Politics of Technology Grant, Intel Research  
2006 William V. Power Award, University of California at Berkeley  
2006 2nd place, CRV/Haas School of Business Venture Capital Competition  
2003-2004 Watson Fellowship, Thomas J. Watson Foundation  
2003 Phi Beta Kappa, Wesleyan University  
2003 Senior Prize in Computer Science, Wesleyan University  
2000 Johnston Prize in Physics, Wesleyan University

## Experience

### Research Experience

- **Thomas J. Watson Foundation, Fellow, 2003-2004**  
Project: “Outreach or Evangelism? Integrating Technology into Unexposed Communities”  
Awarded \$22,000 grant to conduct qualitative research on access to information and communication technologies in China, South Africa, Cape Verde, Argentina, Brazil and Chile.
- **Microsoft Research, Intern, 2002**  
MyLifeBits project, supervised by Jim Gray and Gordon Bell  
Designed and prototyped next-generation databases that supported hyperlinks, annotations, reports, saved queries, pivoting, clustering, and fast search.

- **Harvard Institutes of Medicine**, *Research Associate*, 2001  
Supervised by Roderick Jensen and Steve Gullans  
Developed quantitative methods for analyzing DNA microarrays. Conducted exploratory data analysis using techniques from machine learning and computational statistics.

## Work Experience

- **HOTorNOT.com**, *Founding Engineer*, 2005-2006  
One of four engineers that built and ran the world's largest dating website for 18-24 year olds. The site served over five million registered users and was run on >200 servers spread across two colocations. The company was sold in 2008 for \$20 million.
- **Uncanny Systems**, *Cofounder/Researcher*, 2000  
Developed algorithms to analyze financial data and predict market fluctuations using agent-based unsupervised learning.

## Publications

### Journal Articles

- Gordon, GJ, RV Jensen, LL Hsiao, SR Gullans, **JE Blumenstock**, WG Richards, MT Jaklitsch, DJ Sugarbaker, and R Bueno (2003). "Using Gene Expression Ratios to Predict Outcome Among Patients With Mesothelioma", *Journal of the National Cancer Institute*, 95(8), 598-605
- Gordon, GJ, RV Jensen, LL Hsiao, SR Gullans, **JE Blumenstock**, S Ramaswamy, WG Richards, DJ Sugarbaker, and R Bueno (2002). "Translation of Microarray Data Into Clinically Relevant Cancer Diagnostic Tests Using Gene Expression Ratios in Lung Cancer and Mesothelioma", *Cancer Research*, 62(17), 4963-4967
- Fryer, RM, J Randall, T Yoshida, LL Hsiao, **JE Blumenstock**, KE Jensen, T Dimofte, RV Jensen, and SR Gullans (2002). "Global Analysis of Gene Expression: Methods, Interpretation, and Pitfalls", *Experimental Nephrology*, 10(2), 64-74
- Hsiao, LL, RV Jensen, T Yoshida, KE Clark, **JE Blumenstock**, and SR Gullans (2002). "Correcting for Signal Errors in the Analysis of Microarray Data", *Biotechniques*, 32 (2), 330-337

### Refereed Conference Proceedings

- **Blumenstock, JE** (2008). "Size Matters: Word Count as a Measure of Quality on Wikipedia", *Proceedings of the 17th International Conference on the World Wide Web*, ACM Press, 2008 (forthcoming).
- Gemmell, J, R Lueder, **JE Blumenstock**, E Solomon, and G Bell (2003). "Telephone, Television and Radio in the Home of the Future", *Internet and Multimedia Systems and Applications*, 2003, 263-268

### Invited Presentations, Technical Reports, and Magazine Articles

- **Blumenstock, JE** (2008). "Automatically Assessing the Quality of Wikipedia Articles", University of California at Berkeley School of Information Technical Report #2008-021.
- **Blumenstock, JE** (2007). "Named Entity Recognition in Biographical Markup", ECAI/PNC Joint Meeting on Area Studies, October 2007.

- **Blumenstock, JE** (2004). “Closing the Gap: Solutions to South Africa’s Digital Divide”, *The Big Issue*. 81(8), 22-23
- **Blumenstock, JE** (2003). “Mining Clusters for Knowledge: Finding Algorithm-Independent Groups in Microarray Data”, Undergraduate thesis, Wesleyan University

## Other Skills

Spoken Languages: Close to fluent in Spanish and Portuguese; conversational in Mandarin.

Computer Languages: Expert programmer in C#/C++, R, Java, PHP, SQL, AJAX/DHTML; Proficient in MatLab, Stata, VBA, Python, ML, S-plus, shell scripts.

Systems: Expert in Windows & Linux systems and administration, with experience maintaining large networks of >200 webservers running linux, apache, mysql, etc.

Internet/Engineering: Expert web designer and web engineer; experience building and running websites serving millions of daily pageviews.