

# Josh Blumenstock

jblumenstock@gmail.com

102 South Hall, School of Information, University of California at Berkeley, Berkeley CA 94720

---

## Education

<b>University of California at Berkeley</b> , Berkeley, CA Pursuing PhD, School of Information	<i>Present</i>
<b>Wesleyan University</b> , Middletown, CT B.A., Computer Science: <i>High Departmental Honors</i> B.A., Physics: <i>High Departmental Honors</i> Honors Thesis: "Mining Clusters for Knowledge: Finding Algorithm-Independent Groups in Microarray Data"	May 2003 Major GPA: 4.0/4.0 Cumulative GPA: 3.9/4.0

---

## Honors and Awards

<i>Berkeley Fellowship</i> , <b>University of California at Berkeley</b>	2006-2010
<i>Power Award</i> , <b>University of California at Berkeley</b>	2006
<i>Watson Fellowship</i> , <b>Thomas J. Watson Foundation</b>	2003-2004
<i>Phi Beta Kappa</i> , <b>Wesleyan University</b>	2003
<i>Senior Prize in Computer Science</i> , <b>Wesleyan University</b>	2003
<i>Johnston Prize in Physics</i> , <b>Wesleyan University</b>	2000

---

## Work/Research Experience

<i>Founding Engineer</i> , <b>HOTorNOT.com</b> , Berkeley, CA • Principal engineer and lead product manager for dating website with over 5M users.	2005-2006
<i>Fellow</i> , <b>Thomas J. Watson Foundation</b> , Providence, RI • Received \$22k grant to research emerging technologies in China, South Africa, Cape Verde, Argentina, Brazil and Chile.	2003-2004
<i>Intern</i> , <b>Microsoft Research</b> , Bay Area Research Center, CA • Worked in Gordon Bell and Jim Gray's lab on MyLifeBits, a database with annotations, pivoting, clustering, and search.	2002
<i>Research Associate</i> , <b>Harvard Institute of Medicine</b> , Boston, MA • Developed statistical techniques to diagnose and classify subtypes of lung cancer and other human disease.	2001
<i>Co-founder/Researcher</i> , <b>Uncanny Systems</b> , Lake Pleasant, MA • Developed algorithms and an application to analyze financial data using agent-based unsupervised learning.	2000

---

## Publications and Presentations

- Blumenstock JE, "Size Matters: Word Count as a Measure of Quality on Wikipedia." **Proceedings of the 17<sup>th</sup> International Conference on the World Wide Web**. ACM Press, 2008 (forthcoming).
- Blumenstock JE, "Automatically Assessing the Quality of Wikipedia Articles." UCB iSchool Tech. Report. 2/2008.
- "Named Entity Recognition in Biographical Markup." *ECAI/PNC Joint Meeting on Area Studies*. October 2007.
- Blumenstock, JE. "Closing the Gap: Solutions to South Africa's Digital Divide" **The Big Issue**. 81:8 (2004): 22-23
- Gemmell J, Lueder R, Blumenstock JE, Solomon E, Bell G. "Telephone, Television and Radio in the Home of the Future" **Internet and Multimedia Systems and Applications**. (2003): 263-268
- Gordon GJ, Jensen RV, Hsiao LL, Gullans SR, Blumenstock JE, ..., Sugarbaker DJ, Bueno R. "Using Gene Expression Ratios to Predict Outcome Among Patients With Mesothelioma." **Journal of the Natl Cancer Inst**. 95:8 (2003): 598-605
- Gordon GJ, Jensen RV, Hsiao LL, Gullans SR, Blumenstock JE, Ramaswamy S, Richards WG, Sugarbaker DJ, Bueno R. "Translation of Microarray Data Into Clinically Relevant Cancer Diagnostic Tests Using Gene Expression Ratios in Lung Cancer and Mesothelioma" **Cancer Research**. 62:17 (2002): 4963-4967
- Fryer RM, Randall J, Yoshida T, Hsiao LL, Blumenstock JE, Jensen KE, Dimofte T, Jensen RV, Gullans SR. "Global Analysis of Gene Expression: Methods, Interpretation, and Pitfalls." **Experimental Nephrology**. 10:2 (2002): 64-74
- Hsiao LL, Jensen RV, Yoshida T, Clark KE, Blumenstock JE, Gullans SR. "Correcting for Signal Errors in the Analysis of Microarray Data." **Biotechniques**. 32:2 (Feb. 2002): 330-337

---

## Skills

- Conversationally fluent in Spanish and Portuguese; conversational in Chinese (Mandarin).
- Research experience in machine learning, data mining, computational statistics.
- Fluent in C#/C++, PHP, VBA, MatLab, SQL. Admin experience on multi-tiered web architecture w/200+ servers.