

# Bingjun Sun

Ph.D. in Computer Science and Engineering  
187 Acalanes Dr APT 15  
Sunnyvale, CA 94086

Email: sunbingjun@hotmail.com  
Phone: (814)360-2567  
Homepage: <http://www.cse.psu.edu/~bsun/>

## Research and Development Background

---

- Intelligent System, Information Retrieval, and Database
- Data Mining and Machine Learning, Text and Graph Mining
- Ubiquitous and Mobil Computing, Location-bases Service, Social Networks.
- Computational Advertising
- Multi-Agent Systems and Artificial Intelligence, Cheminformatics

## Education

---

### Ph.D. in Computer Science and Engineering, Minor in Statistics,

The Pennsylvania State University, University Park, PA, Dec. 2008. (Major GPA: 3.97/4.0)  
Dissertation: Mining, Indexing, and Search Approaches to Entity and Graph Information Retrieval for Chemoinformatics.  
Advisor: C. Lee Giles and Prasenjit Mitra.

### MS in Computer Science and Engineering,

The Pennsylvania State University, University Park, PA, Dec. 2003. (Major GPA: 3.97/4.0)  
Thesis: Energy-Efficient Scheduling Algorithms of Object Retrieval on Indexed Parallel Broadcast Channel.  
Advisor: Ali R. Hurson.

### MA in Architecture, Minor in Urban Planning,

Tsinghua University, Beijing, China, July 2000. (Ranking: 1/44)

### BA in Architecture, Minor in Urban Planning,

Tsinghua University, Beijing, China, July 1997. (Summa Cum Laude, Ranking: 4/80)

## Technical Skills

---

**Programming Languages:** C, C++, Java, Perl, SQL, XML, HTML, VRML, Javascript, Pig Script.

**Platforms:** Windows/DOS, Linux, Solaris Unix, MAC.

**Applications and Tools:** Grid computing, MatLab, Lucene, SAS, Photoshop, 3DMax, AutoCAD, MS Office.

## Research and Professional Experiences

---

**TeleNav Inc.**, Senior Software Engineer, Tech Lead

March 2009- present

- Location-based search, advertising, and social network. System design and development. Data mining and modeling.

**Yahoo! Labs.** *Search and Advertising Science*, Scientist

June 2008- Feb. 2009

- Developed a system of new ad impression forecasting for non-guaranteed content match ads. Responsibility includes both data mining and system issues, such as large size log data sampling and processing, multiple parallel index building using grid computing, off-line model training, and online forecasting using a multithread http server.

**Google Inc.** *NewYork Office*, Software Engineer Intern

May 2007-Aug. 2007

- Webpage classification for blog search to identify blogs by training classifiers to filter spam blogs and non-blogs. Worked on both training and testing processes using grid computing of mapreduce. Achieved high precision and recall about 96% for English webpages and about 99% for Chinese webpages.

**Pennsylvania State University.** *Intelligent Information Systems Research Lab, College of IST*, Jan. 2006-May 2008

- Graph mining, indexing, and search with the application of cheminformatics. Handled various graph query models. Developed algorithms of frequent subgraph pattern mining, canonical labeling, subgraph isomorphism, and maximum common subgraph mining. Proposed a probabilistic model for near-optimal subgraph feature

selection. Proposed a new learnable fast graph kernel to rank retrieved graphs by measuring similarity of graphs. Developed a graph search system. Papers are under review.

- Chemical entity tagging, indexing, and search with full document indexing and search, a part of an NSF-funded project of an academic digital library, Chem<sub>x</sub>Seer, the Chemical Cyber-infrastructure. Developed a system of the chemical entity search with document search, including crawling, entity tagging, indexing, and search. Proposed and developed algorithms of chemical entity tagging based on hierarchical conditional random fields, independent frequent subsequence mining, and unsupervised text segmentation. Papers are accepted by WWW 2007 and WWW 2008.
- Blog data mining on prediction of blogging behavior using temporal and social network analysis. Paper is accepted by ICDM 2007.
- Proposed and developed algorithms of multi-document topic segmentation and alignment and shared topic detection using mutual information, term weights, and dynamic programming for optimization. Papers are accepted by SIGIR 2007 and CIKM 2006.
- Automatic scoring project on Biology reports by semi-structured text classification in the Lion Writer project. Developed algorithms of adaboost of decision trees and feature extraction.

**Pennsylvania State University.** *Intelligent Agents Lab, College of IST,* Jan. 2004-Jan. 2006

- Multi-agent system framework based on the computational collaborative Recognition-Primed Decision model.
- Combat simulation for Army Research Lab. Spatial and temporal limited resource allocation under uncertainty. Papers are accepted by AAMAS 2006, KIMAS 2005, and BRIMS 2005.

**Pennsylvania State University.** *Global Information Systems Research Group, Dept of CSE,* Aug. 2002- Dec. 2003

- Project on mobile database. Proposed and implemented fast algorithms of energy-efficient object retrieval from parallel broadcast channels on mobile database using wireless networking. Paper is accepted by ICPP 2004.
- Research on wireless mobile ad hoc networks and security.

**Pennsylvania State University.** *CAD & Immersive Environments Lab, Dept of Architecture,* Aug. 2000–July 2002

- Human computer interaction, virtual reality, and web-based collaborative virtual design studio.

**Tsinghua University,** Beijing, China. *School of Architecture* Jan. 1992–Jun. 2000

- Computer graphic applications, 3d modeling, rendering, and animation, AutoLisp programming in AutoCAD, and web development and design.
- Collaborative urban design project between Tsinghua Univ. and MIT.
- Outstanding undergraduate diploma project on urban planning. Statistic analysis and GIS application.
- Project of the New Campus of National College of Fine Art and the New Campus of Shanghai University.
- Project among top 3 in Cities with Soul Sino-British Students' Architecture and Urban Design Competition 1998 as the team leader of 6 team members.
- Project among top 1% (top 3 in Asia) in ACSA/OTIS Elevator International Student Design Competition 1996.
- Project won the grand prize (top 1) in Shanghai International Design Competition on Residence 1996.

## Publications

---

### Refereed Journal Papers in Progress:

1. Bingjun Sun, Prasenjit Mitra, and C. Lee Giles, "Towards Chemical-Entity-Aware Information Retrieval," under review. 2009.
2. Bingjun Sun, Prasenjit Mitra, and C. Lee Giles, "Data Mining Approaches to Efficient Graph Search," in progress. 2009.
3. Ali R. Hurson, Bingjun Sun, John Hannan, and Y. Jiao, "Power-Aware Scheduling of Data Retrieval from Indexed Parallel Broadcast Channels," under review. 2009.
4. Xiacong Fan, Michael McNeese, Bingjun Sun, Timothy Hanratty, Laurel Allender, and John Yen, "Human-Agent Collaboration for Time Stressed Multi-Context Decision Makings", under review. 2009.

### Conference Papers under Review:

5. Bingjun Sun, Prasenjit Mitra, and C. Lee Giles, "Irredundant Informative Subgraph Mining for Graph Search on the Web," under review. 2009.

6. Bingjun Sun, Prasenjit Mitra, and C. Lee Giles, "Learning to Rank Graphs for Online Similarity Search," under review. 2009.

#### Conference, Workshop, and other Papers:

7. Bingjun Sun, Qingzhao Tan, Prasenjit Mitra, and C. Lee Giles, "Extraction and Search of Chemical Formulae in Text Documents on the Web," in Proceedings of the International World Wide Web Conference (**WWW07**). 2007. **Best Student Paper Nominee**. [acceptance rate: 14%; best paper & best student paper nominee rate: 1%]
8. Bingjun Sun, Prasenjit Mitra, and C. Lee Giles, "Mining, Indexing, and Searching for Textual Chemical Molecule Information on the Web," in Proceedings of the International World Wide Web Conference (**WWW08**). 2008. [acceptance rate: 12%]
9. Bingjun Sun, Prasenjit Mitra, Hongyuan Zha, C. Lee Giles, and John Yen, "Topic Segmentation with Shared Topic Detection and Alignment of Multiple Documents," in Proceedings of the International ACM SIGIR Conference on Research & Development in Information Retrieval (**SIGIR07**). 2007. [acceptance rate: 17%]
10. Bi Chen, Qiankun Zhao, Bingjun Sun, and Prasenjit Mitra, "Temporal and Social Network Based Blogging Behavior Prediction In BlogSpace," in Proceedings of the International Conference on Data Mining (**ICDM07**). 2007. [acceptance rate: 12%]
11. Xiaocong Fan, Bingjun Sun, Shuang Sun, Michael McNeese, John Yen, "RPD-enabled Agents Teaming with Human for MultiContext Decision Making," in Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems (**AAMAS06**). 2006. [acceptance rate: 17%]
12. Bingjun Sun, Ding Zhou, Hongyuan Zha, and John Yen, "Multi-Task Text Segmentation and Alignment Based on Weighted Mutual Information," in Proceedings of the Conference on Information and Knowledge Management (**CIKM06**). 2006. [acceptance rate: 25%]
13. Xiaocong Fan, Rui Wang, Bingjun Sun, Shuang Sun, and John Yen. "Multi-Agent Information Dependence," in Proceedings of the International Conference Integration of Knowledge Intensive Multi-Agent Systems (**KIMAS05**). 2005.
14. Xiaocong Fan, Shuang Sun, Bingjun Sun, Guruprasad Airy, Michael McNeese, John Yen, "Collaborative RPD-enabled Agents Assisting - The Three-Block Challenge in C<sup>2</sup>CUT," in Proceedings of the Conference on Behavior Representation in Modeling and Simulation (**BRIMS05**). 2005.
15. Bingjun Sun, Ali R. Hurson, and John Hannan, "Energy-Efficient Scheduling Algorithms of Object Retrieval on Indexed Parallel Broadcast Channels," in Proceedings of the International Conference on Parallel Processing (**ICPP04**). 2004. [acceptance rate: 34%]
16. Prasenjit Mitra, C. Lee Giles, Bingjun Sun, Ying Liu, and Anuj R. Jaiswal, "Scientific Data and Document Processing in ChemXSeer," in AAAI Spring Symposium 2008.
17. Prasenjit Mitra, C. Lee Giles, Bingjun Sun, Ying Liu, "Information Extraction from Scientific Documents in ChemXSeer," in AAAI/SSS Workshop on Semantic Scientific Knowledge Integration (SSKI08). 2008.
18. K. T. Mueller, P. Mitra, C. L. Giles, B. J. Garrison, J. D. Kubicki, S. L. Brantley, Bingjun Sun, Y. Liu, W. J. Brouwer, S. Nangia, and J. Z. Bandstra. "ChemXSeer: Cyber-tools for researchers in environmental chemistry," Division of Chemical Information for the 235th American Chemical Society National Meeting, 2008.
19. C. L. Giles, P. Mitra, K. Muller, J. Kubicki, B. Garrison, J. Z. Wang, B. Sun, Y. Liu, Q. Tan, L. Bolelli, X. Lu, A. Jaiswal, K. Bai, I. Councill, W. Brouwer, J. Fernandez, and J. Bandstra, "ChemXSeer: An eScience Web Search Engine and Repository for Chemistry," in iConference08. 2008.
20. C. L. Giles, P. Mitra, K. Mueller, J. Kubicki, B. Garrison, J. Z. Wang, B. Sun, L. Bolelli, X. Lu, Y. Liu, I. Councill, W. Brower, Q. Tan, A. Jaiswal, J. Kubicki, B. Garrison and J. Bandstra, "ChemXSeer: An eChemistry Web Search Engine and Repository," in NSF Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation (NGDM07). 2007.
21. C. Lee Giles, Prasenjit Mitra, Levent Bolelli, Xiaonan Lu, Ying Liu, Anuj Jaiswal, Kun Bai, Bingjun Sun, James Z. Wang, Karl Mueller, William Brouwer, James Kubicki, Barbara Garrison, Joel Bandstra, "ChemXSeer: An eChemistry Web Search Engine and Repository," in the Microsoft eScience Workshop. 2007.

22. Prasenjit Mitra, C. Lee Giles, Bingjun Sun, Ying Liu, "ChemXSeer: A Digital Library and Data Repository for Chemical Kinetics," in CIKM07 Workshop on CyberInfrastructure: Information Management in eScience (CIMS07). 2007.
23. Bolelli, L., Lu, X., Liu, Y., Jaiswal, A., Bai, K., Sun, B., Council, I., Mitra, P., Wang, J.Z., Mueller, K., Kubicki, J., Garrison, B., Bandstra J., Giles, C.L. "ChemXSeer: A Chemistry Web Portal for Scientific Literature and Datasets," Open Repositories Conference, San Antonio, Texas, 2007.
24. Fu Li, Bingjun Sun, "Specialization of Social Issues – Providing Rentable Residence for the Floating Population". (Chinese). World Architecture. 1997 (5).

## Teaching Experiences

---

- Pennsylvania State University. *Department of Computer Science and Engineering* Jan. 2003–May. 2003  
Teaching assistant and grader of Logical Design of Digital Systems.
- Pennsylvania State University. *School of Architecture* Aug. 2000–May. 2002  
Instructor of Computer Aided Design Labs.  
Teaching assistant of Introduction to Architecture, Materials and Building Construction.
- Tsinghua University, Beijing, China. *School of Architecture* Sep. 1998–Jan. 1999  
Instructor of the course of Architecture Design.

## Professional Services and Activities

---

- PC member and invited reviewer:  
Special Issue of Machine Learning Journal on Mining and Learning with Graphs 2009.  
ACM Workshop on Social Network Mining and Analysis (SNAKDD) 2009.  
Special Issue on Data Mining for Social Network Data (DMSND. Annals of Information Systems) 2009.  
International Conference on Computational Aspects of Social Networks (CASoN) 2009.  
ACM Workshop on Social Network Mining and Analysis (SNAKDD) 2008.
- Peer reviewer:  
EDBT 2009.  
SDM 2008, SIGIR 2008, KDD 2008, CIKM 2008, WIDM 2008, ICDM 2008.  
WWW 2007, SIGIR 2007, ICDL 2007, AAAI 2007, ICML 2007, ICDM 2007, WIDM 2007.
- Talks:  
Ebay Research Lab, January 2009.  
IBM Almaden Research Center, December 2008.  
Pennsylvania Stat University, Dissertation Defense, September 2008.  
Missouri University of Science and Technology, Department of Computer Science, March 2008.  
Yahoo! Search and Advertising Science, March 2008.  
Pennsylvania Stat University, IST Graduate Symposium, February 2008.  
Google, Summer Intern Poster Session, August 2007.  
Pennsylvania Stat University, Dissertation Proposal, March 2007.  
The WWW conference, May 2007.  
The Conference on Information and Knowledge Management, November 2006.

## Honors

---

- Best Student Paper Nominee and Student travel award of WWW2007 2007
- Tsinghua University Outstanding Graduate (top 2%) 2000
- Tsinghua University Graduated with Highest Honors (top 5%) 1997
- Tsinghua University Society & Club Activities Awards 1996–1999
- Tsinghua University *Nitianzeng* Fellowship, *12.9* Fellowship, *Guanghua* Fellowships 1995–1999
- Tsinghua University Outstanding Undergraduate 1994–1996
- 1st Class Prize in Physics, Olympiad of High Schools, Sichuan Province 1991
- 2nd Class Prize in Mathematics, Olympiad of High Schools, Sichuan Province 1991