

# Xiao Zhang

Department of Computer Science and Engineering  
The Pennsylvania State University  
University Park, PA 16802

Phone: (814) 308-4830  
E-mail: [xiazhang@cse.psu.edu](mailto:xiazhang@cse.psu.edu)  
Homepage: [www.cse.psu.edu/~xiazhang](http://www.cse.psu.edu/~xiazhang)

## Highlights:

- Strong background in Geo-spatial Information Extraction, Machine Learning, Text Mining and Databases
- Reviewer for two journals: Sensors and Journal of Pattern Recognition Research
- Internship experiences in top technology companies: Microsoft Research Asia, Facebook and Google
- Excellent Oral and Written English communication and presentation skills

## Education:

- Ph.D. Candidate, ABD (Aug. 2006 – present)  
Department of Computer Science & Engineering, the Pennsylvania State University  
GPA: 3.74 / 4.0  
**Estimated Graduation Time: Summer, 2011**
- Bachelor's Degree (Aug. 2002 – July. 2006)  
Department of Computer Science and Engineering  
Harbin Institute of Technology, Harbin, China  
GPA: 89.3 / 100 Rank: top 5 / 280

## Research Interests:

Data Mining and Machine Learning, Geo-Spatial Information Extraction, Location-Based Services, Databases

## Academia Activities:

Reviewer of Sensors journal  
Reviewer of Journal of Pattern Recognition Research (JPRR)

## Industrial Experiences:

- **Google** (May 18<sup>th</sup> 2010 – Aug. 27<sup>th</sup>, 2010)  
Software engineer intern in Search Infrastructure Team, developing the clustering system for search quality user feedbacks, including data pre-processing, feature extraction, result analysis and parameter tuning.
- **Facebook** (May 18<sup>th</sup> 2009 – Aug. 7<sup>th</sup>, 2009)  
Software engineer intern in Search Team, developing type filters and location filters for stream search results. Code in production.
- **Microsoft Research Asia** (May. 19<sup>th</sup> 2008 – Aug. 19<sup>th</sup>, 2008)  
Research intern in Web Search and Data Mining group, doing research in traffic condition prediction

## Publications:

- Xiao Zhang, Prasenjit Mitra, Alexander Klippel, Alan MacEachren, Automatic Extraction of Destinations, Origins and Route Parts from Human Generated Route Directions, In *proceedings of 6<sup>th</sup> international conference on Geographic Information Science (GIScience 2010)*, 2010
- Xiao Zhang, Prasenjit Mitra, Learning Topical Transition Probabilities in Click Through Data with Regression Models, *WebDB 2010*
- Xiao Zhang, Prasenjit Mitra, Sen Xu, Anuj R. Jaiswal, Alex Klippel, Alan MacEachren, Extracting Route Directions from Web Pages, *WebDB 2009*
- Xiao Zhang, Wang-Chien Lee, Prasenjit Mitra and Baihua Zheng, Processing Transitive Nearest Neighbor Queries in Multi-Channel Access Environments. In *Proceedings of Extending Database Technology (EDBT) 2008*, a prestige and tier-1 conference in Database Technology and **acceptance rate around 17%**.
- Alan M. MacEachren, Anuj Jaiswal, Anthony C. Robinson, Scott Pezanowski, Alexander Savelyev, Prasenjit Mitra, Xiao Zhang and Justine Blanford, SensePlace2: GeoTwitter Analytics Support for Situational Awareness. In *IEEE Conference on Visual Analytics Science and Technology (IEEE VAST) 2011*, to appear.
- Xu, S., Jaiswal, A., Zhang, X., Klippel, A., Mitra, P. and MacEachren, A. 2010: From Data Collection to Analysis - Exploring Regional Linguistic Variation in Route Directions by Spatially-Stratified Web Sampling. In *Proceedings of the Workshop on Computational Models of Spatial Language Interpretation at Spatial Cognition 2010*
- Xu, S., Klippel, A., MacEachren, A.M., Mitra, P., Turton, I., Zhang, X. and Jaiswal, A. 2010: Exploring regional variation in spatial language - a case study on spatial orientation with spatially stratified web-sampled documents. In *Rapp, D.N., editor, Spatial Cognition 2010. Poster Presentations*
- Alexander Klippel, Alan MacEachren, Prasenjit Mitra, Ian Turton, Xiao Zhang, Anuj Jaiswal, Kean Soon, Jared Oyler and Rui Li, Geographic Analysis of Linguistically Encoded Movement Patterns - A Contextualized Perspective. In *GIScience 2008*

## Research Experiences:

- **Penn State University** (Sep. 2007 – present)  
Research Assistant in GeoCAM project (Geographic Contextualization for Accounts of Movement, funded by NGA). Doing research in geo-spatial information extraction, geo-spatial disambiguation, and text data mining. Building system for extraction and interpretation of movement patterns and landmarks in text documents. Designed the core information extraction and text mining components of GeoCAM system.  
Designed and implemented the back-end (text processing and information extraction module) of Notice to Mariners visualization tool.
- **Penn State University** (Aug. 2006 – Sep. 2007)  
Research Assistant. Doing research in mobile/wireless databases and location based services. Focusing on query processing and optimization of transitive nearest neighbor (TNN) queries. Developing new distance metrics and algorithm for query processing and a novel optimization approach. Conducting comprehensive experiments for evaluation.
- **Harbin Institute of Technology** (Oct, 2005 – July, 2006)  
Doing research on  $k$ -anonymity and privacy preservation. Implementing an anonymizer using bottom-up, depth-first approach to traverse the lattice to find  $k$ -anonymization strategy for a table.

## Technical Skills:

- Programming: C, C++, JAVA, Matlab, SQL, PHP, C#, UNIX Shell, VC, HTML

- Tools: MALLET, WEKA, PostgreSQL, PostGIS, OpenStreetMap
- Platform: Windows, UNIX, Linux
- Communication: Passed the PSU Oral English Efficiency Test with highest score

**Teaching Experiences:**

- **Penn State University** (Jun. 2007 – Aug. 2007)  
Instructor and Teaching Assistant for C++ and Matlab
- **Penn State University** (Aug. 2007 – Dec. 2007)  
Teaching Assistant for Computer Networks