

## Ke Zhai

Tel no.: (240) 460-9082  
zhaikedavy@gmail.com

### EDUCATION

Ph.D. in Computer Science, Department of Computer Science  
**University of Maryland, College Park, MD, Anticipant Date: Dec 2014**  
GPA: 3.9/4.0. Supervisor: Jordan Boyd-Graber and Jimmy Lin (co-supervised)  
Research topic: Large Scale Inference for Probabilistic Bayesian Models

M.S. in Computer Science, Department of Computer Science  
**University of Maryland, College Park, MD, 2011**  
GPA: 3.9/4.0. Supervisor: Jordan Boyd-Graber and Jimmy Lin (co-supervised)  
Research topic: Variational Bayesian Inference of Latent Dirichlet Allocation in MapReduce  
Master scholarly paper: Using Variational Inference and MapReduce to Scale Topic Modeling

B.Eng. in Computer Engineering with first class honor, School of Computer Engineering  
**Nanyang Technological University, Singapore, 2009**  
GPA: 4.65/5.0. Supervisor: Wee Keong Ng  
Research topic: Privacy-Preserving Data Mining  
Undergraduate thesis: An Embedded Caching Framework for Privacy-Preserving Data Mining

### EMPLOYMENT

**Graduate Research Assistant**, Department of Computer Science, University of Maryland, College Park, MD, Sep 2010 - present

- Academic Advisor: Jordan Boyd-Graber
- Research Interest: Machine Learning, Non-parametric Bayesian Learning, Cloud Computing
- Design and implement online variational inference for latent Dirichlet allocation with infinite vocabulary.
- Design and implement variational inference for latent Dirichlet allocation in MapReduce.
- Design and implement variational inference for Indian buffet process in MapReduce.
- Enable variational inference for synthetic topic modelling system in MapReduce.

**Research Intern** Yahoo! Labs, New York City, NY, Jun 2014 - Aug 2014

- Mentor: Prasad Chalasani
- Large scale unsupervised nonparametric models for user behavior analysis.

**Research Intern** Microsoft Research, Redmond, WA, May 2013 - Aug 2013

- Mentor: Jason D. Williams
- Design and implement three models in discovering latent structure in dialogues.
- Achieve comparably well results against many other models on real datasets.

**Software Engineering Intern**, comScore, Inc., Reston, VA, May 2010 - Aug 2010

- Implement data transfer and formatter block for new deployed Hadoop distributed file system.
- Research and develop cookie deletion and prediction system for comScore.

**Graduate Teaching Assistant**, Department of Computer Science, University of Maryland, College Park, MD, Sep 2009 - May 2010

- CMSC 131: Object-Oriented Programming I, Fall 2009.
- CMSC 422: Introduction to Artificial Intelligence, Spring 2010.

**Software Engineering Industrial Attachment**, StarHub, Inc., Singapore, Jan 2008 - Jun 2008

- Study on customer service system and client feedback process.
- Develop data collection and analysing software for Customer Service.
- Develop high-efficiency data processing system.
- Database development and administration.
- Design and develop reliable database system in MySQL for Customer Service.
- Automate data processing and analysing system with Microsoft Office and VBA.

## PUBLICATIONS (\* indicates equal contributor)

**Ke Zhai**, Jordan Boyd-Graber and Shay B. Cohen. “Online Adaptor Grammars with Hybrid Inference”. *Transaction of the Association for Computational Linguistics (TACL)*, to appear.

**Ke Zhai**, and Jason D. Williams. “Discovering Latent Structure in Task-Oriented Dialogues”. *Annual Meeting of the Association for Computational Linguistics (ACL)*, Jun 2014.

**Ke Zhai\***, Yuening Hu\*, Vladimir Edelman, and Jordan Boyd-Graber. “Polylingual Tree-Based Topic Models for Translation Domain Adaptation”. *Annual Meeting of the Association for Computational Linguistics (ACL)*, Jun 2014.

**Ke Zhai**, and Jordan Boyd-Graber. “Online Latent Dirichlet Allocation with Infinite Vocabulary”. *International Conference on Machine Learning (ICML)*, Jun 2013.

**Ke Zhai\***, Yuening Hu\*, Jordan Boyd-Graber, and Sinead Williamson. “Modeling Images using Transformed Indian Buffet Processes”. *International Conference on Machine Learning (ICML)*, Jun 2012.

**Ke Zhai**, Jordan Boyd-Graber, Nima Asadi, and Mohamad Alkhouja. “Mr. LDA: A Flexible Large Scale Topic Modeling Package using Variational Inference in MapReduce”. *ACM International Conference on World Wide Web (WWW)*, Apr 2012.

**Ke Zhai**, Wee Keong Ng, Andre Ricardo Herianto and Shuguo Han. “Speeding Up Secure Computations via Embedded Caching”. *Proceedings of SIAM International Conference on Data Mining (SDM)*, Apr 2009.

## PROFESSIONAL CONTRIBUTION

Activate contributor for MapReduce library Cloud<sup>9</sup> and Hadoop toolkit Ivory.  
Online released code-base on GitHub, <https://github.com/kzhai>.

- MapReduce latent Dirichlet allocation (Variational Bayesian inference, with extension to informed prior and polylingual LDA).
- Latent Dirichlet allocation (Gibbs sampling, variational Bayesian inference and online version).
- Non-parametric Bayesian models (Indian buffet process, hierarchical Dirichlet process and infinite Gaussian mixture model).

## PROFESSIONAL SKILLS

Java, Python, Matlab, MapReduce, Hadoop, Spark, SQL, C++/C, VBA  
Linux, Windows, Mac OS X  
Microsoft Office, Eclipse, NetBeans, Emacs, Photoshop