

## Benyah Shaparenko

benyah@cs.cornell.edu  
http://www.benyah.net

4130 Upson Hall  
Cornell University  
Ithaca, NY 14853

### RESEARCH INTERESTS

- ◇ Data Mining, Machine Learning, Text Mining, Temporal Data Mining
- ◇ Sponsored Search, Information Retrieval, Multimedia Retrieval

### EDUCATION

**Cornell University**, Ithaca, NY 2003-Present  
Expected Aug. 2009, Ph.D. in Computer Science with a minor in Chinese Literature  
Thesis: Information Genealogy: Influential Idea Flows in Text (Adviser: Thorsten Joachims)  
**Pennsylvania State University**, University Park, PA 1999-2003  
B.S. in Computer Science with a minor in Mathematics  
Graduated with Highest Distinction, with Honors in Computer Science  
Thesis: Private State in Object-Oriented Programming Languages (Adviser: John Hannan)

### SELECTED WORK EXPERIENCE

**Research Assistant**, Cornell 2005-Present  
Research text mining methods for detecting original and influential content in documents  
**Search and Advertising Sciences Intern**, Yahoo!, Santa Clara, CA Summer 2008  
Constructed text features to predict when users click on sponsored search advertisements  
**Software Engineering Intern**, Google, New York, NY Summer 2007  
Designed methods to generate short, human-readable, descriptive document titles  
**Visiting Researcher**, National Taiwan Univ., Taipei, Taiwan Summer 2006

### PROFESSIONAL ACTIVITIES

- ◇ Program Committee, ECML/PKDD 2008

### SELECTED AWARDS

- ◇ TA Achievement Award, Computer Science Dept., Cornell University 2005
- ◇ Student Marshal, Computer Science, College of Engineering, Penn State University 2003
- ◇ National Science Foundation Research Experience for Undergraduates 2002
- ◇ Dean's List, each semester, Penn State University 1999 – 2003

### SELECTED PUBLICATIONS

- ◇ Benyah Shaparenko and Thorsten Joachims. **Identifying the Original Contribution of a Document via Language Modeling**. Paper in submission. Poster accepted at SIGIR-2009.
- ◇ Benyah Shaparenko and Thorsten Joachims. **Information Genealogy: Uncovering the Flow of Ideas in Non-hyperlinked Document Databases**. In *Proceedings of KDD-07*, pp. 619-628, 2007.
- ◇ Benyah Shaparenko, Rich Caruana, Johannes Gehrke, and Thorsten Joachims. Identifying Temporal Patterns and Key Players in Document Collections. In *Proceedings of the IEEE ICDM Workshop on Temporal Data Mining (TDM-05)*, pp. 165-174, Houston, TX, USA, 2005.
- ◇ A. M. MacEachren, G. Cai, R. Sharma, I. Rauschert, I. Brewer, L. Bolelli, B. Shaparenko, S. Fuhrmann, and H. Wang, Enabling Collaborative Geoinformation Access and Decision-Making through a Natural, Multimodal Interface, *International Journal of Geographical Information Science*, 19(3), pp. 293-317, 2005.
- ◇ B. Shaparenko and J. Hannan. Private State in Object-Oriented Programming Languages. Honors Thesis. The Pennsylvania State University, 2003.

### SELECTED OTHER ACTIVITIES

- ◇ **Co-Founder** and **Co-President**, Chinese Language Club, Cornell 2007-2008
- ◇ GPSA Finance Commission, Cornell 2007-2008