

## Benyah Shaparenko

*Email:* benyah@cs.cornell.edu      *Web Site:* <http://www.benyah.net>  
*Permanent:* (717) 679-0438      *School:* (607) 255-3009  
7745 Lancaster Ave.      4130 Upson Hall  
Myerstown, PA 17067      Ithaca, NY 14853

### RESEARCH INTERESTS

- ◇ Clustering, Machine Learning, Text Classification
- ◇ Data Mining, Text Mining, Temporal Data Mining
- ◇ Natural Language Processing

### EDUCATION

**Cornell University**, Ithaca, NY      August 2003 – Present  
Major: Computer Science, Minor: Chinese Literature  
Advisor: Thorsten Joachims

**Pennsylvania State University**, University Park, PA      August 1999 – May 2003  
B.S., Major: Computer Science, Minor: Mathematics  
Graduated with Highest Distinction, with Honors in Computer Science  
Advisor: John Hannan

### WORK EXPERIENCE

**Research Assistant**, Computer Science Department, Cornell University, Ithaca, NY  
Supervisor: Thorsten Joachims      May 2005 – Present

- ◇ Perform temporal text mining research on influential document and author analysis
- ◇ Research the stability of clustering algorithms when resampling from data
- ◇ Write reports and maintain a log of research experiments, discussions, and ideas
- ◇ Attend conferences, present publications, and review paper submissions
- ◇ Previously conducted research on text classification methods for spam filtering

**Teaching Assistant**, Computer Science Department, Cornell University, Ithaca, NY  
Supervisors: Graeme Bailey, Daisy Fan, Fred Schneider      August 2003 – May 2005  
Courses: COM S 100M/100J: Introduction to Computer Programming (Matlab/Java)  
COM S 230: Intermediate Design and Programming for the Web  
COM S 513: Systems Security

- ◇ Taught “section,” a group of 20-30 students, in typical classroom format
- ◇ Substituted to teach regular class lecture when professor was traveling
- ◇ Conducted lab sessions and office hours, helping students both individually and in groups
- ◇ Designed, proctored, and graded homework projects and exam questions

**SAT Instructor**, Kaplan, State College, PA  
Supervisor: Stacey Dorang      June 2003 – July 2003

- ◇ Taught Kaplan’s SAT preparation for 9th, 10th, and 11th grade high school students in the Summer Study at Penn State program
- ◇ Proctored exams, held office hours, and handled student situations

**Research Assistant**, GEOVISTA Studio, Geography Department, Penn State, University Park, PA  
Supervisor: Alan MacEachren      October 2001 – July 2003

- ◇ Created a reader and writer for ESRI shapefiles
- ◇ Improved human-computer multi-modal interface system in summer 2002
- ◇ Sped up system response time by optimizations (e.g. directly loading PNG files to display)
- ◇ Helped design XML strings for natural language utterances for a knowledge base to interpret

- ◇ Developed SuperWaba client for Dell Axim X5 handheld devices and Java server using sockets

**Computer Science Intern**, Spotts, Stevens, and McCoy Engineering Consultants, Reading, PA  
 Supervisor: Dave Quinter June 2001 – August 2001

- ◇ Developed database software to track building permits in MS Access
- ◇ Web-deployed the permit application using ASP with C++ components
- ◇ Designed an internal database for the PA SWAP project

**Research Assistant**, Physics Department, Penn State Berks, Reading, PA  
 Supervisor: Daniel Litvin June 2000 – June 2001

- ◇ Developed crystallography software using MFC in Visual C++
- ◇ Researched physics and math theories in crystallography

**Tutor**, The Learning Center, Penn State Berks, Reading, PA  
 Supervisor: Anna Donahue August 1999 – May 2001

- ◇ Tutored computer science, math, and physics

## AWARDS

- ◇ TA Achievement Award, Computer Science Department, Cornell University 2005
- ◇ Student Marshal, Computer Science, College of Engineering, Penn State University 2003
- ◇ National Science Foundation Research Experience for Undergraduates 2002
- ◇ Evan Pugh Scholar Award, Penn State University 2002 – 2003
- ◇ President Sparks Award, Penn State University 2001
- ◇ Chris Mader Memorial Scholarship, Penn State University 2001 – 2003
- ◇ Lockheed Martin Engineering Scholars Award 2001 – 2002
- ◇ President's Freshman Award, Penn State University 2000
- ◇ PHEAA SciTech Scholarship 2000 – 2002
- ◇ Harold W. and Jeane Perkins Scholarship, Penn State Berks 2000 – 2001
- ◇ Boscov Honors Excellence Award, Penn State Berks 2000 – 2001
- ◇ Boscov Honors Scholar Award, Penn State Berks 2001
- ◇ Boscov Honors Scholar Program Endowment 1999 – 2001
- ◇ Dean's List, each semester, Penn State University 1999 – 2003
- ◇ Paul Morrow Endowed Scholarship, College of Engineering, Penn State University 1999 – 2002
- ◇ National Deans' List 1999 – 2002
- ◇ All-American Scholar 1999 – 2001
- ◇ Melville C. Barr Scholarship, Penn State Berks 1999 – 2000
- ◇ Presidential Ces Scholarship, Penn State Berks 1999 – 2000

## PUBLICATIONS

- ◇ 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: Algorithms, Theory and Applications (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.
- ◇ J. Privratska, B. Shaparenko, D. Litvin, and V. Janovec. Magnetic Point Group Symmetries of Spontaneously Polarized and/or Magnetized Domain Walls, *Ferroelectrics*, 269, pp. 39-44, 2002.
- ◇ B. Shaparenko, J. Schlessman, and D. B. Litvin. Domain Average Engineering in Ferroics,

*Ferroelectrics*, 269, pp. 9-14, 2002.

LANGUAGE PROFICIENCY

- ◇ English, native speaker
- ◇ Mandarin Chinese (reading, writing, and speaking), fluent (4 years of college-level courses)

COMPUTER SKILLS

- ◇ Microsoft Windows, RedHat Linux
- ◇ Java, Perl, C++, Matlab, L<sup>A</sup>T<sub>E</sub>X, SQL, XML, HTML