

## BIOGRAPHICAL SKETCH

**Thomas W. Finley**

PhD Student in Computer Science  
Cornell University, Ithaca, New York

### Education

<u>Institution</u>	<u>Area</u>	<u>Degree</u>	<u>Year</u>
Cornell University	Computer Science (Current GPA : 4.104) Committee members: Thorsten Joachims (advisor), Rich Caruana, Johannes Gehrke, David Williamson	Ph.D. Candidate	2008?
Duke University	Computer Science (Overall GPA : 3.832) Honors advisor Susan Rodger	B.S.	2002
Duke University	Mathematics Honors advisor William Allard	B.S.	2002
Duke University	Economics	Minor	2002

### Appointments

<u>Title</u>	<u>Employer</u>	<u>Period</u>
Software Engineering Intern	Google Inc.	May '07 - Aug. '07
Research Intern	Yahoo! Research Labs	May '05 - Aug. '05
Graduate Research Assistant	Cornell University	Dec. '03 - Dec. '04, Aug. '05 - Dec. '05, May '06 - Dec '06, Jan. '08' - present
Graduate Teaching Assistant	Cornell University	Aug. '03 - Dec. '03, Jan. '05 - May '05, Jan. '06 - May '06, Jan. '07 - Dec. '07
Undergrad Researcher (JFLAP)	Duke University	Jun. '02 - Aug. '03
Undergrad Researcher (JAWAA)	Duke University	Jun. '01 - Aug. '01
Undergrad Researcher (PRUV)	Duke University	Jul. '01 - Aug. '01, Jan. '02 - Apr. '02
Undergrad Teaching Assistant	Duke University	Sep. '00 - May. '02

### Awards

International Conference on Machine Learning Distinguished Distinguished Student Paper Award (2005)

Elected to the Duke Chapter of Phi Beta Kappa (Spring 2003)

Computing Research Association Outstanding Undergraduate Honorable Mention (2003)  
Graduation with High Distinction in Computer Science (for JFLAP)  
Graduation with High Distinction in Mathematics (for PRUV work on Myrinet)  
Practical Research for Undergraduates using VIGRE (PRUV) Fellowship  
ACM Programming Contest Mid-Atlantic Region: 10th of 139 teams (2001), 11th of 136 (2002), 4th of 146 and qualification for the world finals (2003)  
Dean's List at Duke (6 semesters), with Distinction 4 of 6

## Publications

Thomas Finley, Thorsten Joachims, "Supervised k-Means Clustering," (*In submission*), 2008.

Thomas Finley, Thorsten Joachims, "Training Structural SVMs when Exact Inference is Intractable," *International Conference on Machine Learning (ICML) (to appear)*, 2008.

Thorsten Joachims, Thomas Finley, and Chun-Nan J. Yu, "Cutting-plane training of structural SVMs." *Machine Learning Journal (to appear)*, 2008.

Thomas Finley, Thorsten Joachims, "Parameter Learning for Loopy Markov Random Fields with Structural Support Vector Machines," *ICML Workshop on Constrained Optimization and Structured Output Spaces*, 2007.

Yisong Yue, Thomas Finley, Filip Radlinski, Thorsten Joachims, "A Support Vector Method for Optimizing Average Precision," *Proceedings of the Conference on Research and Development in Information Retrieval (SIGIR)*, 2007.

Susan Rodger and Thomas Finley, "JFLAP - An Interactive Formal Languages and Automata Package," ISBN 0763738344, Jones and Bartlett, published 2/27/06, 2006.

Susan H. Rodger, Bart Bressler, Thomas Finley, and Stephen Reading, "Turning Automata Theory into a Hands-on Course," *Thirty-seventh SIGCSE Technical Symposium on Computer Science Education*, 2006.

Thomas Finley and Thorsten Joachims, "Supervised Clustering with Support Vector Machines," *Proceedings of the International Conference on Machine Learning (ICML)*, 2005.

Ryan Cavalcante, Thomas Finley and Susan H. Rodger, "A Visual and Interactive Automata Theory Course with JFLAP 4.0," *Thirty-fifth SIGCSE Technical Symposium on Computer Science Education*, p. 140-144, 2004.

Ayonike Akingbade, Thomas Finley, Diana Jackson, Pretesh Patel and Susan H. Rodger, "JAWAA: Easy Web-Based Animation from CS 0 to Advanced CS Courses," *Thirty-fourth SIGCSE Technical Symposium on Computer Science Education*, p. 162-166, 2003.

## Projects & Software

**Supervised K-means/Spectral Clustering with SVMs** – Devised a supervised  $k$ -means / spectral clustering algorithm, where we train partitionings of sets to learn a similarity measure. This differs from previous work in that it parameterizes with respect to the  $k$ -means/spectral clustering objective. (2007-2008). (with Thorsten Joachims)

**Approximate Inference Structured Learning** – As part of my thesis work, I theoretically and empirically analyzed the SVM learner for functions with complex outputs in settings which required approximate algorithms for inference, using loopy Markov random fields as a common model that requires approximate inference. (2006-2008). (with Thorsten Joachims)

**PyGLPK** – Wrote PyGLPK, a Python module that encapsulates the functionality of the GNU Linear Programming Kit (GLPK), a toolkit for solving linear programming problems. The goal of PyGLPK is to give one access to all documented functionality of GLPK within a Pythonic (that is, non-SWIG) Python module. (2007). <http://www.cs.cornell.edu/~tomf/pyglpk/>

**PyGraphcut** – Wrote a Python module that encapsulates the functionality of some of the graph cut code written by Vladimir Kolmogorov so that it could be easily exploited by Python code. My goal in writing the code is to support my own research, which required as a subcomponent an engine to solve quadratic pseudo-Boolean optimization problems. (2007). <http://www.cs.cornell.edu/~tomf/pyglpk/>

**Rank Retrieval Optimization** – Contributed to an algorithm for parameterizing ranked retrieval systems (a search engine is the most obvious instance of this) so that mean average precision, a common ranked-retrieval performance measure, is maximized over a training set directly in a computationally efficient manner. (2006-2007) (with Yisong Yue, Filip Radlinski, and Thorsten Joachims)

**Supervised Correlation Clustering with SVMs** – Devised a supervised correlation clustering algorithm where we train over partitionings of sets to learn a similarity measure. This utilizes an existing framework  $SVM^{struct}$  which learn functions with complex output spaces with SVMs. We learn to map a set of items to a partitioning over that set. Our current applications include document noun-phrase coreference, and clustering of related news stories. (2004-2005). (with Thorsten Joachims)

**$SVM^{python}$  and  $SVM^{python} 2$**  – Developed  $SVM^{python}$ , an extension to the existing  $SVM^{struct}$  package.  $SVM^{struct}$  allows users to write interface functions in C to enable machine learning of functions with complex outputs.  $SVM^{python}$  embeds Python, allowing users to write the functions in Python instead of C. This enables more rapid prototyping of new ideas. (2005-present) <http://www.cs.cornell.edu/~tomf/svmpython2/>

**Automata Learning Package** – Developed JFLAP 4.0, a new version of a visual interactive software tool for teaching automata theory. JFLAP lets users create structures related to complexity theory (automata, regular expressions, grammars, L-systems)

and perform guided procedures on structures that correct mistakes students make. (2002-03). (with Susan Rodger) <http://www.cs.duke.edu/~rodger/tools/jflap/>

**JFLAP User Manual** – Coauthored a user manual for JFLAP 4.0 (I wrote 6 of the 11 chapters). This manual is intended as a supplementary text for an automata theory course, providing discussion of relevant of automata theory in the context of JFLAP, and exercises making use of JFLAP. (2003-2005, published 2006) (with Susan Rodger)

**JFLAP** – Authored the Java Formal Language and Automata Package version 4.0, a widely used educational interactive visualization tool for automata theory and formal languages, including visualizations and algorithms covering automata (finite state, push-down, Turing machines), grammars, regular expressions, and L-systems. I was the sole developer past summer 2002. (2002-2005) (with Susan Rodger and Ryan Cavalcante)

**Mac Freeware** – Developed Fob, freeware software for Mac OS X that lets one set count-down timers. The purpose was to teach myself Objective-C and the Cocoa API. Fob received three-and-a-half mice (July 2004) and a later version four mice (November 2004) from Macworld. (2003-2004) <http://homepage.mac.com/tfinley/LeakyPuppy/>

**Animation Tool** – Developed a JAWAA editor for easily defining JAWAA animations in a GUI. JAWAA is an animation language primarily used to visualize algorithms. (2001). (with Susan Rodger) <http://www.cs.duke.edu/csed/jawaa2/>

**Network Routing** – Researched algorithms to efficiently route packets over a Myrinet, a high speed network for a network of workstations as part of my Practical Research for Undergraduates using VIGRE (PRUV) fellowship. (2001). (with William Allard)

**ACM Contest** – Participated in the ACM International Collegiate Programming Contest for Duke for three years including competing in the world finals in 2003. Was also vice president of student chapter of the ACM at Duke in 2001.

## Skills

**Languages** – C, C++, Objective-C, Python, Java, MATLAB, Maple, MIPS assembler, PERL, SML, Scheme, Lisp, Prolog, Applescript.

**Important Course Work** – [at Duke] CPS106 (prog. languages), CPS124 (graphics), CPS140 (complexity theory), CPS150 (numerical analysis), CPS170 (method. in AI), CPS230 (grad. algorithm analysis), MATH121 (abstract algebra), MATH124 (combinatorics), MATH135 (probability), MATH224 (grad. sci. computing), [at Cornell] CS611 (prog. languages), CS632 (adv. databases), CS672 (adv. AI), CS678 (adv. topics in machine learning), INFO640 (adv. human-computer interfaces), INFO650 (language and technology)